PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07K 14/435, 14/47, 16/00, C12N 5/16, 15/12, 15/85 | A1

(11) International Publication Number: WO 95/30693

(43) International Publication Date: 16 November 1995 (16.11.95)

(21) International Application Number: PCT/US95/05741

(22) International Filing Date: 8 May 1995 (08.05.95)

(30) Priority Data:
08/239,228 6 May 1994 (06.05.94) US

(60) Parent Application or Grant
(63) Related by Continuation
US 08/239,228 (CIP)
Filed on 6 May 1994 (06.05.94)

(71) Applicant *(for all designated States except US)*: FRED HUTCHINSON CANCER RESEARCH CENTER [US/US]; 1124 Columbia Street, Seattle, WA 98104 (US).

(71) Applicant *(for US only)*: WEINTRAUB, Nancy (executrix for the deceased inventor) [US/US]; 1223-16 Avenue E., Seattle, WA 98112 (US).

(72) Inventor: WEINTRAUB, Harold, M. (deceased).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: LEE, Jacqueline, E. [US/US]; 922 S. Ogden, Denver, CO 80209-4426 (US). HOLLENBERG, Stanley, M. [US/US]; 6413 SW Roundtree Court, Portland, OR 97219 (US). TAPSCOTT, Stephen, J. [US/US]; 3419 S. Dose Terrace, Seattle, WA 98144 (US).

(74) Agent: BRODERICK, Thomas, F.; Christensen, O'Connor, Johnson & Kindness, Suite 2800, 1420 Fifth Avenue, Seattle, WA 98101 (US).

(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: NEUROGENIC DIFFERENTIATION (NEUROD) GENES AND PROTEINS

(57) Abstract

An isolated nucleic acid molecule which comprises at least 15 nucleotides and which hybridizes under stringent conditions with a *neuro D* HLH domain selected from among nucleotides 577-696 of SEQ ID NO:1, nucleotides 376-495 of SEQ ID NO:3, nucleotides 405-524 of SEQ ID NO:8, nucleotides 273-392 of SEQ ID NO:10, and complements thereof.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

Neurogenic Differentiation (NeuroD) Genes and Proteins

This invention was made with government support under grant CA42506 awarded by the National Institutes of Health. The government has certain rights in the invention.

This application is a continuation-in-part of U.S. Serial No. 08/239,238, filed May 6, 1994.

Field of the Invention

The invention relates to molecular biology and in particular to genes and proteins involved in vertebrate neural development.

Background of the Invention

There are currently several examples of transcription regulatory proteins sharing a basic helix-loop-helix (bHLH) secondary structure. bHLH proteins form homodimeric and heterodimeric complexes binding DNA in the 5' regulatory regions of genes controlling expression. Among the bHLH proteins, mammalian MyoD and *Drosophila* AS-C are presently thought to play developmental roles in muscle development and in sensory organ development, respectively. Both proteins are thought to exert their effects by binding 5' regulatory nucleotide sequences in genes that seem specifically determinative of cellular differentiation and fate. However, the specific developmental roles of the genes affected by MyoD and AS-C remain largely unknown, as are the molecular details of the developmental pathways regulated by these genes. The presently disclosed NeuroD represents a new sub-family of bHLH proteins and is implicated in vertebrate neuronal development.

Neural tissues and endocrine tissues do not regenerate. Damage is permanent. Paralysis, loss of vision or hearing, and hormonal insufficiency are also permanent.

Tumors in neural and endocrine tissues can also be very difficult to treat because of the toxic side effects that conventional chemotherapeutic drugs may have on nervous tissues. The medical community and public would greatly benefit from the availability of agents active in triggering differentiation in neuroectodermal stem cells. Such neuronal differentiating agents could be used for construction of test cell lines, assays for identifying candidate therapeutic agents capable of inducing regeneration of neuronal and endocrine tissues, gene therapy, and differentiation of tumor cells.

Summary of the Invention

Mammalian and amphibian NeuroD proteins were identified, and polynucleotide molecules encoding NeuroD were isolated and sequenced. *NeuroD* encodes a protein that is a distinctive member of the bHLH family. In addition, the present invention provides a family of NeuroD proteins that share a highly conserved HLH region. In the neurula stage of the mouse embryo (e10), *neuroD* is highly expressed in the neurogenic derivatives of neural crest cells, the cranial and dorsal root ganglia, and postmitotic cells in the central nervous system (CNS). During mouse development, *neuroD* is expressed transiently and concomitant with neuronal differentiation in differentiating neurons in sensory organs such as in nasal epithelium and retina. In *Xenopus* embryos ectopic expression of *neuroD* in non-neuronal cells induced formation of neurons.

A representative nucleotide sequence of murine *neuroD* is shown in SEQ ID NO:1. The HLH coding domain of murine *neuroD* resides between nucleotides 577 and 696 in SEQ ID NO:1. The deduced amino acid sequence of murine NeuroD is shown in SEQ ID NO:2. There is a highly conserved region following the helix-2 domain from amino acid 150 through amino acid 199 of SEQ ID NO:2 that is not shared by other bHLH proteins.

A representative nucleotide sequence of *Xenopus neuroD* is shown in SEQ ID NO:3. The HLH coding domain of *Xenopus neuroD* resides between nucleotides 376 and 495 in SEQ ID NO:3. The deduced amino acid sequence of murine NeuroD is shown in SEQ ID NO:4. There is a highly conserved region following the helix-2 domain from amino acid 157 through amino acid 199 of SEQ ID NO:4 that is not shared by other bHLH proteins.

Human *neuroD* sequences are also disclosed. Representative nucleotide and deduced amino acid sequences of the human NeuroD family of are shown in SEQ ID NOS:8-11. The disclosed human clones, 9F1 and 14B1, have an identical HLH motif: amino acid residues 117-156 in SEQ ID NO:9 and residues 91-130 in SEQ ID NO:11.

Brief Description of the Drawings

FIGURE 1 schematically depicts the domain structure of the murine and *Xenopus* NeuroD bHLH proteins.

Detailed Description of the Preferred Embodiment

Tissue-specific bHLH proteins that regulate early neuroectodermal differentiation were discovered using expression cloning and screening assays designed to identify possible bHLH proteins capable of interacting with the protein product of the *Drosophila daughterless* gene. These proteins belong to a family of proteins that share conserved residues in the HLH region.

NeuroD is a member of a novel protein family and is found to be transiently expressed in differentiating neurons during embryogenesis. Its expression is also detected in adult brain, in the granule layer of hippocampus and cerebellum. NeuroD contains the basic helix-loop-helix (bHLH) domain structure that has been implicated in the binding of bHLH proteins to upstream recognition sequences and activation of downstream target genes. The present invention provides representative NeuroD proteins, which include the murine NeuroD protein of SEQ ID NO:2 and the amphibian NeuroD protein of SEQ ID NO:4. Based on homology with other bHLH proteins, the bHLH domain for the murine NeuroD protein is predicted to reside between amino acids 102 and 155 of SEQ ID NO:2, and between amino acids 101 and 157 of SEQ ID NO:4 for the amphibian NeuroD protein. As detailed below, the present invention provides the identification of the human *neuroD* and, in addition, provides an unexpected homologous gene of the same family based on the almost identical sequence across the HLH domain shared between the two human genes at the amino acid level. NeuroD proteins are transcriptional activators that control transcription of downstream target genes that cause neuronal progenitors to differentiate into mature neurons. As discussed in more detail below, NeuroD proteins are expressed in differentiating neurons and are capable of causing the conversion of non-neuronal cells into neurons. The present invention encompasses NeuroD variants that, for example, are modified in a manner that results in a NeuroD protein capable of binding to its recognition site, but unable to activate downstream genes. NeuroD proteins encompass proteins retrieved from naturally occurring materials and closely related, functionally similar proteins retrieved by antisera specific to NeuroD, and recombinantly expressed proteins encoded by genetic materials (DNA, RNA, cDNA) retrieved on the basis of their similarity to the unique regions in the neuroD family of genes.

The present invention discloses representative isolated and purified polynucleotide molecules encoding proteins of the NeuroD family. Representative polynucleotide molecules encoding NeuroD include the sequences presented in SEQ ID NOS:1, 3, 8, and 10. Polynucleotide molecules encoding NeuroD include those sequences resulting in minor genetic polymorphisms, differences between species, those that contain amino acid substitutions, additions, and/or deletions.

In some instances, one may employ such changes in the sequence of recombinant NeuroD to substantially decrease or even increase the biological activity of NeuroD, depending on the intended use of the preparation. Such changes may also be directed towards endogenous *neuroD* sequences using, for example, gene therapy methods to alter the gene product.

The NeuroD proteins of the present invention are capable of inducing the expression of neuronal-specific genes, such as N-CAM, β-tubulin, and Xen-1, neurofilament M (NF-M), Xen-2, tanabin-1, shaker-1, and frog HSCL, in a frog embryo. As described below, NeuroD activity may be detected when NeuroD is ectopically expressed in frog oocytes following, for example, injection of *neuroD* RNA into one of the two cells in a two-cell stage *Xenopus* embryo, and monitoring expression of neuronal-specific genes in the injected as compared to un-injected side of the embryo by immunochemistry or *in situ* hybridization.

"Over-expression" means an increased level of NeuroD protein or *neuroD* transcripts in a recombinant transformed host cell relative to the level of protein or transcripts in the parental cell from which the host cell is derived.

As noted above, the present invention provides isolated and purified polynucleotide molecules encoding NeuroD and other members of the NeuroD family. The disclosed sequences may be used to identify and isolate *neuroD* polynucleotide molecules from suitable host cells such as canine, ovine, bovine, caprine, lagomorph, or avian. In particular, the nucleotide sequences encoding the HLH region may be used to identify poylnycleotide molecules encoding other proteins of the NeuroD family. Complementary DNA molecules encoding NeuroD family members may be obtained by constructing a cDNA library mRNA from, for example, fetal brain. DNA molecules encoding NeuroD family members may be isolated from such a library using the disclosed sequences in standard hybridization techniques (e.g., Sambrook et al., ibid., and Bothwell, Yancopoulos and Alt, ibid.) or by amplification of sequences using polymerase chain reaction (PCR) amplification (e.g, Loh et al. *Science* 243: 217-222, 1989; Frohman et al., *Proc. Natl. Acad. Sci. USA* 85: 8998-9002, 1988; and Erlich (ed.), *PCR Technology: Principles and Applications for DNA*

Amplification, Stockton Press, 1989; which are incorporated by reference herein in their entirety). In a similar manner, genomic DNA encoding NeuroD may be obtained using probes designed from the sequences disclosed herein. Suitable probes for use in identifying *neuroD* sequences may be obtained from *neuroD*-specific sequences that are highly conserved regions between mammalian and amphibian *neuroD* coding sequences. Primers, for example, from the region encoding the approximately 40 residues following the helix-2 domain are suitable for use in designing PCR primers. Alternatively, oligonucleotides containing specific DNA sequences from a human *neuroD* coding region may be used within the described methods to identify human *neuroD* genomic and cDNA clones. Upstream regulatory regions of *neuroD* may be obtained using the same methods. Suitable PCR primers are between 7-50 nucleotides in length, more preferably between 15 and 25 nucleotides in length. Alternatively, *neuroD* polynucleotide molecules may be isolated using standard hybridization techniques with probes of at least about 7 nucleotides in length and up to and including the full coding sequence. Southern analysis of mouse genomic DNA probed with the murine *neuroD* cDNA under stringent conditions showed the presence of only one gene, suggesting that under stringent conditions bHLH genes from other protein families will not be identified. Other members of the *neuroD* family can be identified using degenerate oligonucleotides based on the sequences disclosed herein for PCR amplification or by hybridization at moderate stringency.

A DNA molecule encoding NeuroD is inserted into a suitable expression vector, which is in turn used to transfect or transform a suitable host cell. Suitable expression vectors for use in carrying out the present invention comprise a promoter capable of directing the transcription of a polynucleotide molecule of interest in a host cell. Representative expression vectors may include both plasmid and/or viral vector sequences. Suitable vectors include retroviral vectors, vaccinia viral vectors, CMV viral vectors, BlueScript® vectors, baculovirus vectors, and the like. Promoters capable of directing the transcription of a cloned gene or cDNA may be inducible or constitutive promoters and include viral and cellular promoters. For expression in mammalian host cells, suitable viral promoters include the immediate early cytomegalovirus promoter (Boshart et al., *Cell* 41: 521-530, 1985) and the SV40 promoter (Subramani et al., *Mol. Cell. Biol.* 1: 854-864, 1981). Suitable cellular promoters for expression of proteins in mammalian host cells include the mouse metallothionien-1 promoter (Palmiter et al., U.S. Patent No. 4,579,821), a mouse Vk promoter (Bergman et al., *Proc. Natl. Acad. Sci.* 81: 7041-7045, 1983; Grant et al. *Nucleic Acid. Res.* 15: 5496, 1987), and tetracycline-responsive promoter (Gossen

and Bujard, *Proc. Natl. Acad. Sci.USA* 89: 5547-5551, 1992 and Pescini et al., *Biochem. Biophys. Res. Comm.* 202: 1664-1667, 1994). Also contained in the expression vectors, typically, is a transcription termination signal located downstream of the coding sequence of interest. Suitable transcription termination signals include the early or late polyadenylation signals from SV40 (Kaufman and Sharp, *Mol. Cell. Biol.* 2:1304-1319, 1982), the polyadenylation signal from the Adenovirus 5 e1B region, and the human growth hormone gene terminator (DeNoto et al., *Nucleic Acid Res.* 9: 3719-3730, 1981). Mammalian cells, for example, may be transfected by a number of methods including calcium phosphate precipitation (Wigler et al., *Cell* 14: 725, 1978; Corsaro and Pearson, *Somatic Cell Genetics* 7: 603, 1981; Graham and Van der Eb, *Virology* 52: 456, 1973); lipofection, microinjection, and electroporation (Neumann et al., *EMBO J.* 1: 8410845, 1982). Mammalian can be transduced with virus such as SV40, CMV, and the like. In the case of viral vectors, cloned DNA molecules may be introduced by infection of susceptible cells with viral particles. Retroviral vectors may be preferred for use in expressing NeuroD in mammalian cells particularly if NeuroD is used for gene therapy (for review, see, Miller et al. *Methods in Enzymology* 217: 581-599, 1994; which is incorporated herein by reference in its entirety). It may be preferable to use a selectable marker to identify cells that contain the cloned DNA. Selectable markers are generally introduced into the cells along with the cloned DNA molecules and include genes that confer resistance to drugs, such as neomycin, hygromycin, and methotrexate. Selectable markers may also complement auxotrophies in the host cell. Yet other selectable markers provide detectable signals, such as beta-galactosidase to identify cells containing the cloned DNA molecules. Selectable markers may be amplifiable. Such amplifiable selectable markers may be used to amplify the number of sequences integrated into the host genome.

As would be evident to one of ordinary skill in the art, the polynucleotide molecules of the present invention may be expressed *Saccharomyces cerevisiae*, filamentous fungi, and *E. coli*. Methods for expressing cloned genes in *Saccharomyces cerevisiae* are generally known in the art (see, "Gene Expression Technology," *Methods in Enzymology*, Vol. 185, Goeddel (ed.), Academic Press, San Diego, CA, 1990, and "Guide to Yeast Genetics and Molecular Biology," *Methods in Enzymology*, Guthrie and Fink (eds.), Academic Press, San Diego, CA, 1991; which are incorporated herein by reference). Filamentous fungi may also be used to express the proteins of the present invention; for example, strains of the fungi *Aspergillus* (McKnight et al., U.S. Patent No. 4,935,349, which is incorporated herein by

reference). Methods for expressing genes and cDNAs in cultured mammalian cells and in *E. coli* is discussed in detail in Sambrook et al. (*Molecular Cloning: A Laboratory Manual, Second Edition,* Cold Spring Harbor, NY, 1989; which is incorporated herein by reference). As would be evident to one skilled in the art, one could express the protein of the instant invention in other host cells such as avian, insect, and plant cells using regulatory sequences, vectors and methods well established in the literature.

The term "capable of hybridizing under stringent conditions" as used herein means that the subject nucleic acid molecules (whether DNA or RNA) anneal to an oligonucleotide of 15 or more contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:8, or SEQ ID NO:10.

NeuroD proteins produced according to the present invention may be purified using a number of established methods such as affinity chromatography using anti-NeuroD antibodies coupled to a solid support. Fusion proteins of antigenic tag and NeuroD can be purified using antibodies to the tag. Additional purification may be achieved using conventional purification means such as liquid chromatography, gradient centrifugation, and gel electrophoresis, among others. Methods of protein purification are known in the art (see generally, Scopes, R., *Protein Purification,* Springer-Verlag, NY, 1982, which is incorporated herein by reference) and may be applied to the purification of recombinant NeuroD described herein.

The choice of hybridization conditions will be evident to one skilled in the art and will generally be guided by the purpose of the hybridization, the type of hybridization (DNA-DNA or DNA-RNA), and the level of desired relatedness between the sequences. Methods for hybridization are well established in the literature; See, for example: Sambrook, ibid.; Hames and Higgins, eds, *Nucleic Acid Hybridization, A Practical Approach*, IRL Press, Washington DC, 1985; Berger and Kimmel, eds, *Methods in Enzymology, Vol. 52, Guide to Molecular Cloning Techniques*, Academic Press Inc., New York, NY, 1987; and Bothwell, Yancopoulos and Alt, eds, Methods for *Cloning and Analysis of Eukaryotic Genes*, Jones and Bartlett Publishers, Boston, MA 1990; which are incorporated by reference herein in their entirety. One of ordinary skill in the art realizes that the stability of nucleic acid duplexes will decrease with an increased number and location of mismatched bases; thus, the stringency of hybridization may be used to maximize or minimize the stability of such duplexes. Hybridization stringency can be altered by: adjusting the temperature of hybridization; adjusting the percentage of helix-destabilizing agents, such as formamide, in the hybridization mix; and adjusting the temperature and salt

concentration of the wash solutions. In general, the stringency of hybridization is adjusted during the post-hybridization washes by varying the salt concentration and/or the temperature. Stringency of hybridization may be reduced by reducing the percentage of formamide in the hybridization solution or by decreasing the temperature of the wash solution. High stringency conditions may involve high temperature hybridization (e.g., 65-68°C in aqueous solution containing 4-6 X SSC, or 42°C in 50% formamide) combined with high temperature (e.g., 5-25°C below the T_m) and a low salt concentration (e.g., 0.1 X SSC). Reduced stringency conditions may involve lower hybridization temperatures (e.g., 35-42°C in 20-50% formamide) with intermediate temperature (e.g., 40-60°C) and washes in a higher salt concentration (e.g., 2-6 X SSC). Moderate stringency conditions, which may involve hybridization at a temperature between 50°C and 55°C and washes in 0.1 X SSC, 0.1% SDS at between 50°C and 55°C, may be used to identify clones encoding members of the NeuroD family.

The invention provides isolated and purified polynucleotide molecules encoding NeuroD capable of hybridizing under stringent conditions an oligonucleotide of 15 or more contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:8, SEQ ID NO:10, and their complementary strands. The subject isolated *neuroD* polynucleotide molecules preferably encode NeuroD proteins that trigger differentiation in ectodermal cells, particularly neuroectodermal stem cells, and in more committed cells of that lineage, for example, epidermal precursor cells. Such *neuroD* expression products typically form heterodimeric bHLH protein complexes that bind in the 5'-regulatory regions of target genes and enhance or suppress transcription of the target gene.

In some instances, cancer cells may contain non-functional NeuroD protein or may contain no NeuroD protein due to genetic mutation or somatic mutations such that these cells fail to differentiate. For cancers of this type, the cancer cells may be treated in a manner to cause the over-expression of wild-type NeuroD protein to force differentiation of the cancer cells.

Antisense *neuroD* nucleotide sequences may be used to block expression of mutant *neuroD* expression in neuronal precursor cells to generate and harvest neuronal stem cells. The use of antisense oligonucleotides and their applications have been reviewed in the literature (see, for example, Mol and Van der Krul, eds., *Antisense Nucleic Acids and Proteins Fundamentals and Applications*, New York, NY, 1992; which is incorporated by reference herein in its entirety). Suitable antisense oligonucleotides are at least 11 nucleotide in length and may include

untranslated (upstream or intron) and associated coding sequences. As will be evident to one skilled in the art, the optimal length of antisense oligonucleotide is its on the strength of the interaction between the antisense oligonucleotide and its complement on the mRNA, the temperature and ionic environment translation takes place, the base sequence of the antisense oligonucleotide, and the presence of secondary and tertiary structure in the mRNA and/or in the antisense oligonucleotide. Suitable target sequences for antisense oligonucleotides include intron-exon junctions (to prevent proper splicing), regions in which DNA/RNA hybrids will prevent transport of mRNA from the nucleus to the cytoplasm, initiation factor binding sites, ribosome binding sites, and sites that interfere with ribosome progression. A particularly preferred target region for antisense oligonucleotide is the 5' untranslated (promoter/enhancer) region of the gene of interest. Antisense oligonucleotides may be prepared by the insertion of a DNA molecule containing the target DNA sequence into a suitable expression vector such that the DNA molecule is inserted downstream of a promoter in a reverse orientation as compared to the gene itself. The expression vector may then be transduced, transformed or transfected into a suitable cell resulting in the expression of antisense oligonucleotides. Alternatively, antisense oligonucleotides may be synthesized using standard manual or automated synthesis techniques. Synthesized oligonucleotides may be introduced into suitable cells by a variety of means including electroporation, calcium phosphate precipitation, or microinjection. The selection of a suitable antisense oligonucleotide administration method will be evident to one skilled in the art. With respect to synthesized oligonucleotides, the stability of antisense oligonucleotide-mRNA hybrids may be increased by the addition of stabilizing agents to the oligonucleotide. Stabilizing agents include intercalating agents that are covalently attached to either or both ends of the oligonucleotide. Oligonucleotides may be made resistant to nucleases by, for example, modifications to the phosphodiester backbone by the introduction of phosphotriesters, phosphonates, phosphorothioates, phosphoroselenoates, phosphoramidates, or phosphorodithioates. Oligonucleotides may also be made nuclease resistant by the synthesis of the oligonucleotides with alpha-anomers of the deoxyribonucleotides.

NeuroD binds to 5' regulatory regions of neurogenic genes that are involved in neuroectodermal differentiation, including development of neural and endocrine tissues. The NeuroD protein alters expression of the subject gene by, for example, down-regulating or up-regulating transcription, or by inducing a change in transcription to an alternative open reading frame. The subject polynucleotide

molecules find a variety of uses, e.g., in preparing oligonucleotide probes, expression vectors, and transformed host cells, as disclosed below in the following Examples.

DNA sequences recognized by NeuroD may be determined using a number of methods known in the literature including immunoprecipitation (Biedenkapp et al, *Nature* 335: 835-837, 1988, Kinzler and Vorgelstein, *Nuc. Acids Res.* 17: 3645-3653, 1989; and Sompayrac and Danna, *Proc. Natl. Acad. Sci. USA* 87: 3274-3278, 1990; which are incorporated by reference herein), protein affinity columns (Oliphant et al., *Mol. Cell. Biol.* 9: 2944-2949, 1989; which is incorporated by reference herein), gel mobility shifts (Blackwell and Weintraub, *Science* 250: 1104-1110, 1990; which is incorporated by reference herein), and Southwestern blots (Keller and Maniatis, *Nuc. Acids Res.* 17:4675-4680, 1991; which is incorporated by reference herein).

One embodiment of the present invention involves the construction of inter-species hybrid NeuroD proteins to facilitate structure-function analyses or to alter NeuroD activity by increasing or decreasing the transcriptional activation of neurogenic genes by NeuroD relative to the wild-type NeuroD. Hybrid proteins of the present invention may contain the replacement of one or more contiguous amino acids of the native NeuroD with the analogous amino acid(s) of NeuroD from another species. Such interspecies hybrid proteins include hybrids having whole or partial domain replacements. As would be evident to one skilled in the art, such hybrid proteins may be obtained using recombinant DNA techniques. Briefly, DNA molecules encoding the hybrid NeuroD proteins of interest are prepared using generally available methods such as PCR mutagenesis, site-directed mutagenesis, and/or restriction digestion and ligation. The hybrid DNA is then inserted into expression vectors and transformed or transfected into suitable host cells. The biological activity may be assessed essentially as described in the assays set forth in more detail in the Examples that follow.

The invention also provides synthetic peptides, recombinantly derived peptides, fusion proteins, and the like. The subject peptides have an amino acid sequence encoded by a nucleic acid which hybridizes under stringent conditions with an oligonucleotide of 15 or more contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:8, or SEQ ID NO:10. Representative amino acid sequences of the subject peptides are disclosed in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:9, and SEQ ID NO:11. The subject peptides find a variety of uses, including preparation of specific antibodies.

As noted above, the invention provides antibodies which bind to NeuroD. The production of non-human antisera or monoclonal antibodies (e.g., murine,

lagormorpha, porcine, equine) is well known and may be accomplished by, for example, immunizing an animal with NeuroD protein or peptides. For the production of monoclonal antibodies, antibody producing cells are obtained from immunized animals, immortalized and screened, or screened first for the production of the antibody that binds to the NeuroD protein or peptides and then immortalized. It may be desirable to transfer the antigen binding regions (i.e., F(ab')2 or hypervariable regions) of non-human antibodies into the framework of a human antibody by recombinant DNA techniques to produce a substantially human molecule. Methods for producing such "humanized" molecules are generally well known and described in, for example, U.S. Patent No. 4,816,397; which is incorporated by reference herein in its entirety. Alternatively, a human monoclonal antibody or portions thereof may be identified by first screening a human B-cell cDNA library for DNA molecules that encode antibodies that specifically bind to NeuroD according to the method generally set forth by Huse et al. (*Science* 246: 1275-1281, 1989, which is incorporated by reference herein in its entirety). The DNA molecule may then be cloned and amplified to obtain sequences that encode the antibody (or binding domain) of the desired specificity.

The invention also provides methods for inducing the expression of genes associated with neuronal phenotype in a cell that does not normally express those genes. Examples of neuronal phenotypes that may be modulated by NeuroD expression include expression of neurotransmitters or neuromodulatory factors. Cells that can be used for the purpose of modulation of gene expression by NeuroD include cells of the neuroectodermal lineage, glial cells, neural crest cells, and epidermal epithelial basal stem cells, and all types of both mesodermal and endodermal lineage cells.

As illustrated in Example 10, the expression of NeuroD protein in stem cells causes redirection of epidermal cell differentiation and induces terminal differentiation into neurons, i.e., instead of epidermal cells. Epithelial basal stem cells (i.e., in skin and mucosal tissues) are one of the few continuously regenerating cell types in an adult mammal. Introduction of the subject nucleotide sequences into an epithelial basal stem cell may be accomplished *in vitro* or *in vivo* using a suitable gene therapy vector delivery system (e.g., a retroviral vector), a microinjection technique (see, for example, Tam, *Basic Life Sciences* 37: 187-194, 1986, which is incorporated by reference herein in its entirety), or a transfection method (e.g., naked or liposome encapsulated DNA or RNA) (see, for example, *Trends in Genetics* 5: 138, 1989; Chen and Okayama, *Biotechniques* 6: 632-638, 1988; Mannino and Gould-Fogerite,

eliminated. Such cells may contain altered *neuroD* coding sequences that result in the expression of a NeuroD protein that is not capable of enhancing, suppressing or activating transcription of the target gene. The subject cell lines and animals find uses in screening for candidate therapeutic agents capable of either substituting for a function performed by NeuroD or correcting the cellular defect caused by a defective NeuroD. Considering the important regulatory role of NeuroD in embryonic development, birth defects may occur from expression of mutant NeuroD proteins, and these defects may be correctable *in utero* or in early post-natal life through the use of compounds identified in screening assays using NeuroD. In addition, *neuroD* polynucleotide molecules may be joined to reporter genes, such as β-galactosidase or luciferase, and inserted into the genome of a suitable embryonic host cell such as an mouse embryonic stem cell by, for example, homologous recombination (for review, see Capecchi, *Trends in Genetics* 5: 70-76, 1989; which is incorporated by reference). Cells expressing NeuroD may then be obtained by subjecting the differentiating embryonic cells to cell sorting, leading to the purification of a population of neuroblasts. Neuroblasts may be useful for studying neuroblast sensitivity to growth factors or chemotherapeutic agents. The neuroblasts may also be used as a source from which to purify specific protein products or gene transcripts. These products may be used for the isolation of growth factors, or for the identification of cell surface markers that can be used to purify stem cell population from a donor for transplantation.

3. Construction of gene transfer vectors (e.g., retroviral vectors, and the like) wherein *neuroD* is inserted into the coding region of the vector under the control of a promoter. *NeuroD* gene therapy may be used to correct traumatic neural injury that has resulted in loss of motor or sensory neural function. For these therapies, gene transfer vectors may either be injected directly at the site of the traumatic injury, or the vectors may be used to construct transformed host cells that are then injected at the site of the traumatic injury. The results disclosed in Example 10 indicate that introduction of *neuroD* induces a non-neuronal cell to become a neuron. This discovery raises for the first time the possibility of using transplantation and/or gene therapy to repair neural defects resulting from traumatic injury. In addition, the discovery of *neuroD* provides the possibility of providing specific gene therapy for the treatment of certain neurological disorders such as Alzheimer's disease, Huntington's disease, and Parkinson's disease, in which a population of neurons have been damaged. Two basic methods of *neuroD* utilization can be envisioned in this regard. In one method, *neuroD* is expressed in existing populations of neurons to modulate

The following examples are offered by way of illustration and not by way of limitation.

EXAMPLE 1

Construction of the embryonic stem cell "179" cDNA library.

A continuous murine embryonic stem cell line (i.e., the ES cell line) having mutant E2A (the putative binding partner of myoD) was used as a cell source to develop a panel of embryonic stem cell tumors. Recombinant ES stem cells were constructed (i.e., using homologous recombination) wherein both alleles of the putative myoD binding partner E2A were replaced with drug-selectable marker genes. ES cells do not make functional E12 or E47 proteins, both of which are E2A gene products. ES cells form subcutaneous tumors in congenic mice (i.e., 129J) that appear to contain representatives of many different embryonal cell types as judged histologically and through the use of RT-PCR gene expression assays. Individual embryonic stem cell tumors were induced in male 129J strain mice by subcutaneous injection of 1×10^7 cells/site. Three weeks later each tumor was harvested and used to prepare an individual sample of RNAs. Following random priming and second strand synthesis the ds-cDNAs were selected based on their size on 0.7% agarose gels and those cDNAs in the range of 400-800 bp were ligated to either Bam HI or Bgl II linkers. (Linkers were used to minimize the possibility that an internal Bam HI site in a cDNA might inadvertently be cut during cloning, leading to an abnormally sized or out-of-frame expression product.) The resultant individual stem cell tumor DNAs were individually ligated into the Bam HI cloning site in the "fl-VP16" 2μ yeast expression vector. This expression vector, fl-VP16, contains the VP16 activation domain of Herpes simplex virus (HSV) located between Hind III (HIII) and Eco RI (RI) sites and under the control of the *Saccharomyces cereviseae* alcohol dehydrogenase promoter; with *LEU2* and Ampicillin-resistance selectable markers. Insertion of a DNA molecule of interest into the Hind III site of the fl-VP16 vector (i.e., 5' to the VP16 nucleotide sequence), or into a Bam HI site (i.e., 3' to the VP16 sequence but 5' to the Eco RI site), results in expression of a VP16 fusion protein having the protein of interest joined in-frame with VP16. The resultant cDNA library was termed the "179-library".

EXAMPLE 2

Identification and cDNA cloning of *neuroD*.

A two-hybrid yeast screening assay was used essentially as described by Fields and Song (Nature 340:245, 1989) and modified as described herein was used to screen the 179-library described in Example 1. Yeast two-hybrid screens are

reviewed as disclosed in Fields and Sternglanz (*Trends in Genetics* 10: 286-292, 1994). The library was screened for cDNAs that interacted with *Lex*A-Da, a fusion protein between the *Drosophila* Da (Daughterless) bHLH domain and the prokaryotic *Lex*A-DNA binding domain. Multimerized *Lex*A binding sites were cloned upstream of two reporter genes, the *HIS3* gene and the β-galactosidase gene. The *S. cereviseae* strain L40 containing a plasmid encoding the *Lex*A-Da fusion protein was transformed with CsCl gradient-purified f1-VP16-179-cDNA library. Transformants were maintained on medium selecting both plasmids (the *Lex*A-Da plasmid and the cDNA library plasmid) for 16 hours before being subjected to histidine selection on plates lacking histidine, leucine, tryptophan, uracil, and lysine. Clones that were HIS^+ were subsequently assayed for the expression of *LacZ*. To eliminate possible non-specific cloning artifacts, plasmids from $HIS^+/LacZ^+$ were isolated and transformed into *S. cereviseae* strain L40 containing a plasmid encoding a *Lex*A-Lamin fusion. Clones that scored positive in the interaction with lamin were discarded. Approximately 400 cDNA clones, which represented 60 different transcripts, were identified as positive in these assays. Twenty-five percent of the original clones were subsequently shown to be known bHLH genes on the basis of their reactivity with specific cDNA probes. One cDNA clone encoding a VP16-fusion protein that interacted with Da but not lamin was identified as unique by sequence analysis. This clone, initially termed *tango*, is now referred to as *neuroD*.

The unique cDNA identified above, VP16-*neuroD*, contained an approximately 450 bp insert that spanned the bHLH region. Sequence analysis showed that the clone contained an insert encoding a complete bHLH amino acid sequence motif that was unique and previously unreported. Further analysis suggested that while the cDNA contained conserved residues common to all members of the bHLH protein family, several residues were unique and made it distinct from previously identified bHLH proteins. The *neuroD* cDNA insert was subcloned as a Bam HI-Not I insert into Bam HI-Not I linearized pBluescript SK^+. The resulting plasmid was designated pSK+1-83.

The *neuroD* insert contained in the VP16-*neuroD* plasmid was used to re-probe a mouse cDNA library prepared from mouse embryos at developmental stage e10.5. Candidate clones were isolated and sequenced essentially as described above. Several clones were isolated. One clone, designated pKS^+ m7a RX, was deposited at the American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD 20852 USA, on May 6, 1994, under accession number 75768. Plasmid pKS^+ m7a RX

contains 1646 bp of murine *neuroD* cDNA as an EcoRI-XhoI insert. The amino acid sequence encoded by the insert begins at amino acid residue +73 and extends to the carboxy-terminus of the NeuroD protein. The plasmid contains about 855 bp of NeuroD coding sequence. (encoding amino acids 73-536).

None of the mouse cDNAs contained the complete 5' coding sequence. To obtain the 5' *neuroD* coding sequence, a mouse strain 129/Sv genomic DNA library was screened with the VP16-*neuroD* plasmid insert (450 bp). Genomic clones were isolated and sequenced and the sequences were aligned with the cDNA sequences. Alignment of the sequence and comparison of the genomic 5' coding sequences with the *Xenopus neuroD* clone (Example 8) confirmed the 5' neuroD coding sequence. The complete *neuroD* coding sequence and deduced amino acid sequence are shown in SEQ ID NOS:1 and 2.

EXAMPLE 3

NeuroD/*neuroD*

bHLH proteins share common structural similarities that include a basic region that binds DNA and an HLH region involved in protein-protein interactions required for the formation of homodimers and heterodimeric complexes. A comparison of the amino acid sequence of the basic region of murine NeuroD (amino acids 102 to 113 of SEQ ID NO:2) with basic regions of other bHLH proteins revealed that murine NeuroD contained all of the conserved residues characteristic among this family of proteins. However, in addition, NeuroD contained several unique residues. These unique amino acid residues were not found in any other known HLH, making NeuroD a distinctive new member of the bHLH family. The NARERNR basic region motif in NeuroD (amino acids 107-113 of SEQ ID NO:2) is also found in the *Drosophila* AS-C protein, a protein thought to be involved in neurogenesis. Similar, but not identical, NARERRR and NERERNR motifs (SEQ ID NOS:5 and 6, respectively) have been found in the *Drosophila* Atonal and MASH (mammalian achaete-scute homolog) proteins, respectively, which are also thought to be involved in neurogenesis. The NARER motif (SEQ ID NO:7) *of neuroD* is shared by other bHLH proteins, and the *Drosophila* Daughterless (Da) and Mammalian E proteins. The basic region of bHLH proteins is important for DNA binding site recognition, and there is homology between NeuroD and other neuro-proteins in this functional region. Within the important dimer-determining HLH region of NeuroD, a low level of homology was recorded with mouse twist protein (i.e., 51% homology) and with MASH (i.e., 46% homology). NeuroD contains several regions of unique peptide sequence within the bHLH domain including the junction sequence (MHG).

EXAMPLE 4

NeuroD is expressed in differentiating neurons during embryonic development.

NeuroD expression was analyzed during embryonic development of mouse embryos using *in situ* hybridization with an antisense *neuroD* single-stranded riboprobe labeled with digoxigenin (Boehringer Mannheim). Briefly, a riboprobe was prepared from plasmid pSK+1-83 using T7 polymerase and digoxigenin-11-UTP for labeling. The hybridized probe was detected using anti-digoxigenin antibody conjugated with alkaline phosphatase. Color development was carried out according to the manufacturer's instruction. Stages of development are commonly expressed as days following copulation and where formation of the vaginal plug is e0.5. The results recorded in the *in situ* hybridization studies were as follows:

In the e9.5 mouse embryo, *neuroD* expression was observed in the developing trigerminal ganglia.

In the e10.5 mouse embryo, a distinctive pattern of *neuroD* expression was observed in all the cranial ganglia (i.e., V-XI) and in dorsal root ganglia (DRG) in the trunk region of the embryo. At this time *neuroD* expression was also observed in the central nervous system in post-mitotic cells in the brain and spinal cord that were undergoing neuronal differentiation. In the spinal cord, the ventral portion of the cord from which the motor neurons arise and differentiate was observed to express *neuroD* at high levels; and expression in the posterior-ventral spinal cord was higher when compared to more mature anterior-ventral spinal cord.

In the e11.5 mouse embryo, the ganglionic expression pattern of *neuroD* observed in e10.5 persisted. Expression in the spinal cord was increased over the level of expression observed in e10.5 embryos, which is consistent with the presence of more differentiating neurons at this stage. At this stage *neuroD* expression is also observed in other sensory organs in which neuronal differentiation occurs, for example, in the nasal epithelium, otic vesicle, and retina of the eye. In both of these organs *neuroD* expression was observed in the region containing differentiating neurons.

In the e14.5 mouse embryo, expression of *neuroD* was observed in cranial ganglia and DRG, but expression of *neuroD* persisted in the neuronal regions of developing sensory organs and the central nervous system (CNS). Thus, *neuroD* expression was observed to be transient during neuronal development.

In summary, expression of *neuroD* in the neurula stage of the embryo (e10), in the neurogenic derivatives of neural crest cells, the cranial and dorsal root ganglia,

and post mitotic cells in the CNS suggests an important possible link between expression and generation of sensory and motor nerves. Expression occurring later in embryonic development in differentiating neurons in the CNS and in sensory organs (i.e., nasal epithelium and retina) also supports a role in development of the CNS and sensory nervous tissue. Since *neuroD* expression is transient, the results suggest that *neuroD* expression is operative as a switch controlling formation of sensory nervous tissue. It is noteworthy that in these studies *neuroD* expression was not observed in embryonic sympathetic and enteric ganglia (also derived from migrating neural crest cells). Overall, the results indicate that *neuroD* plays an important role in neuronal differentiation.

EXAMPLE 5

NeuroD is expressed in neural and brain tumor cells: murine probes identify human *neuroD*.

Given the expression pattern in mouse embryo (Example 4), Northern blots of tumor cell line mRNAs were examined using murine *neuroD* cDNA (Example 2) as a molecular probe. As a first step, cell lines that have the potential for developing into neurons were screened. The D283 human medullablastoma cell line, which expressed many neuronal markers, expressed high levels of *neuroD* by Northern blot analysis. *NeuroD* was also transcribed at various levels by different human neuroblastoma cell lines and in certain rhabdomyosarcoma lines that are capable of converting to neurons. Murine PC12 pheochromacytoma cells and P19 embryocarcinoma cells differentiate into neurons in tissue culture in the presence of appropriate inducers, i.e., nerve growth factor and retinoic acid, respectively. When induced, murine P19 but not PC12 cells expressed *neuroD* transcripts. However, non-induced murine PC12 cells, P19 cells, and control 3T3 fibroblasts did not produce detectable levels of *neuroD* transcripts. Thus, PC12 and P19 cells represent cell types that are potentially useful in screening assays for identifying inducers of *neuroD* expression that may stimulate nerve regeneration and differentiation of neural tumor cells.

EXAMPLE 6

Recombinant cells expressing NeuroD.

Recombinant murine 3T3 fibroblast cells expressing either a myc-tagged murine NeuroD protein or myc-tagged *Xenopus* NeuroD protein were made. The recombinant cells were used as a test system for identifying antibody to NeuroD described below.

Xenopus NeuroD protein was tagged with the antigenic marker Myc to allow the determination of the specificity of anti-NeuroD antibodies to be determined.

Plasmid CS2+MT was used to produce the Myc fusion protein. The CS2+MT vector (Turner and Weintraub, ibid.) contains the simian cytomegalovirus IE94 enhance/promoter (and an SP6 promoter in the 5' untranslated region of the IE94-driven transcript to allow in vitro RNA synthesis) operatively linked to a DNA sequence encoding six copies of the Myc epitope tag (Roth et al, *J. Cell Biol.* 115: 587-596, 1991; which is incorporated herein in its entirety), a polylinker for insertion of coding sequences, and an SV40 late polyadenylation site. CS2-MT was digested with Xho I to linearize the plasmid at the polylinker site downstream of the DNA sequence encoding the *myc* tag. The linearized plasmid was blunt-ended using Klenow and dNTPs. A full length *Xenopus* cDNA clone was digested with Xho I and Eae I and blunt-ended using Klenow and dNTPs, and the 1.245 kb fragment of the *Xenopus neuroD* cDNA was isolated. The *neuroD* fragment and the linearized vector were ligated to form plasmid CS2+MT x1-83.

CS2+MT was digested with Eco RI to linearize the plasmid at the polylinker site downstream of the DNA sequence encoding the *myc* tag. The linearized plasmid was blunt-ended using Klenow and dNTPs and digested with Xho I to obtain a linearized plasmid having an Xho I adhesive end and a blunt end. Plasmid pKS+m7a containing a partial murine *NeuroD* cDNA was digested with Xho I, and the NeuroD containing fragment was blunt-ended and digested with Xba I to obtain the approximately 1.6 kb fragment of the murine *neuroD* cDNA. The *neuroD* fragment and the linearized vector were ligated to form plasmid CS2+MT M1-83(m7a).

Plasmids CS2+MT x1-83 and CS2+MT M1-83(m7a) were each transformed into murine 3T3 fibroblast cells and used as a test system for identifying antibody against NeuroD (Example 7).

EXAMPLE 7

Antibodies to NeuroD.

A recombinant fusion protein of maltose binding protein (MBP) and amino acid residues 70-355 of murine NeuroD was used as an antigen to evoke antibodies in rabbits. Specificity of the resultant antisera was confirmed by immunostaining of the recombinant 3T3 cells described above. Double-immunostaining of the recombinant cells was observed with monoclonal antibodies to Myc (i.e., the control antigenic tag on the transfected DNA) and with rabbit anti-murine NeuroD in combination with anti-rabbit IgG. The specificity of the resultant anti-murine NeuroD sera was investigated further by preparing mouse 3T3 fibroblasts cells transfected with different portions of NeuroD DNA. Specificity seemed to map to the glutamic acid-rich domain (i.e., amino acids 66-73 of SEQ ID NO:2). The anti-murine antisera did not

react with cells transfected with the *myc*-tagged *Xenopus neuroD*. In a similar manner, *Xenopus* NeuroD was used to generate rabbit anti-NeuroD antisera. The antisera was *Xenopus*-specific and did not cross react with cells transfected with *myc*-tagged murine *neuroD*.

EXAMPLE 8

NeuroD is a highly evolutionarily conserved protein: sequence of *Xenopus* NeuroD.

Approximately one million clones from a stage 17 *Xenopus* head library made by Kintner and Melton (*Development* 99: 311, 1987) were screened with the mouse cDNA insert as a probe at low stringency. The hybridization was performed with 50% formamide/4 X SSC at 33°C and washed with 2 X SSC/0.1% SDS at 40°C.

Positive clones were identified and sequenced. Analysis of the *Xenopus neuroD* cDNA sequence (SEQ ID NO:3) revealed that NeuroD is a highly conserved protein between frog and mouse. The deduced amino acid sequences of frog and mouse (SEQ ID NOS:2 and 4) show 96% identity in the bHLH domain (50 of 52 amino acids are identical) and 80% identity in the region that is carboxy-terminal to the bHLH domain (159 of 198 amino acids are identical). The domain structures of murine and *Xenopus* NeuroD are highly homologous with an "acidic" N-terminal domain (i.e., glutamic or aspartic acid rich); a basic region; helix 1, loop, helix 2; and a proline rich C-terminal region. Although the amino terminal regions of murine and *Xenopus* NeuroD differ in amino acid sequence, both retain a glutamic or aspartic acid rich "acidic domain" (amino acids 102 to 113 of SEQ ID NO:2 and amino acids 56 to 79 of SEQ ID NO:4). It is highly likely that the acidic domain constitutes an "activation" domain for the NeuroD protein, in a manner analogous to the activation mechanisms currently understood for other known transcription regulatory factors.

EXAMPLE 9

Neuronal expression of *Xenopus neuroD*.

The expression pattern of *neuroD* in whole mount *Xenopus* embryos was determined using *in situ* hybridization with a single stranded digoxigenin-labeled *Xenopus neuroD* antisense cDNA riboprobe. Embryos were examined at several different stages.

Consistent with the mouse expression pattern, by late stage, all cranial ganglia showed very strong staining patterns. In *Xenopus*, as in other vertebrate organisms, neural crest cells give rise to skeletal components of the head, all ganglia of the peripheral nervous system, and pigment cells. Among these derivatives, the cranial sensory ganglia, which are of mixed crest and placode origin, represent the only group

of cells that express *neuroD*. High levels of *neuroD* expression in the eye were also observed, correlating with active neuronal differentiation in the retina at this stage. Expression is observed in the developing olfactory placodes and otic vesicles, as was seen in mice. The pineal gland also expressed *neuroD*. All of this expression in transient, suggesting that *neuroD* functions during the differentiation process but is not required for maintenance of these differentiated cell types.

As early as stage 14 (i.e., the mid-neurula stage) *neuroD* expression was observed in the cranial neural crest region where trigerminal ganglia differentiate. Primary mechanosensory neurons in the spinal cord, also referred to as Rohon-Beard cells and primary motor neurons, showed *neuroD* expression at this stage.

By stage 24, all of the developing cranial ganglia, trigerminal, facio-acoustic, glosso-pharyngeal, and vagal nervous tissues showed a high level of *neuroD* expression. High levels of expression of *neuroD* was also observed in the eye at this stage. (Note that in *Xenopus* neuronal differentiation in the retina occurs at a much earlier stage than in mice, and *neuroD* expression was correspondingly earlier and stronger in this animal model.)

In summary, in *Xenopus* as in mouse, *neuroD* expression was correlated with sites of neuronal differentiation. The remarkable evolutionary conservation of the pattern of *neuroD* expression in differentiating neurons supports the notion that NeuroD has been evolutionarily conserved both structurally and functionally in these distant classes, which underscores the critical role performed by this protein in embryonic development.

EXAMPLE 10

Ectopic expression of *neuroD* converts non-neuronal cells into neurons.

To further analyze the biological functions of NeuroD, a gain-of-function assay was conducted. In this assay, RNA was microinjected into one of the two cells in a 2-cell stage *Xenopus* embryo, and the effects on later development of neuronal phenotype was evaluated. For these experiments *myc*-tagged *neuroD* transcripts were synthesized *in vitro* using SP6 RNA polymerase. The *myc* tagged-*neuroD* transcripts were microinjected into one of the two cells in a *Xenopus* 2-cell embryo, and the other cell of the embryo served as an internal control. Antibodies to *Xenopus* N-CAM, a neural adhesion molecule, anti-Myc (to detect the exogenous protein), and immunostaining techniques were used to evaluate phenotypic expression of the neuronal marker (and control) gene during the subsequent developmental stages of the microinjected embryos. Remarkably, an evaluation of over 130 embryos that were

injected with *neuroD* RNA showed a striking increase in ectopic expression of N-CAM on the microinjected side of the embryo (i.e., Myc^+), as judged by increased immunostaining. The increased staining was observed in the region from which neural crest cells normally migrate. It is considered likely that ectopic expression (or over-expression) of *neuroD* caused neural crest stem cells to follow a neurogenic cell fate. Outside the neural tube, the ectopic immunostaining was observed in the facio-cranial region and epidermal layer, and in some cases the stained cells were in the ventral region of the embryo far from the neural tube. The immunostained cells not only expressed N-CAM ectopically, but displayed a morphological phenotype of neuronal cells. At high magnification, the N-CAM expressing cells exhibited typical neuronal processes reminiscent of axonal processes.

To confirm that the ectopic N-CAM expression resulted from a direct effect on the presumptive epidermal cells and not from aberrant neural cell migration into the lateral and ventral epidermis, *neuroD* RNA was injected into the top tier of 32-cell stage embryos, in order to target the injection into cells destined to become epidermis. N-CAM staining was observed in the lateral and ventral epidermis without any noticeable effect on the endogenous nervous system, indicating that the staining of N-CAM in the epidermis represents the conversion of epidermal cell fate into neuronal cell fate.

Ectopic generation of neurons by *neuroD* was confirmed with other neural specific markers, such as neural-specific class II β-tubulin (Richter et al., *Proc. Natl. Acad. Sci. USA* 85: 8066, 1988), acetylated alpha-tubulin (Piperno and Fuller, *J. Cell. Biol.* 101: 2085, 1985), tanabin (Hemmati-Brinvanlou et al., *Neuron* 9: 417, 1992), neurofilament(NF)-M (Szaro et al., *J. Comp. Neurol.* 273: 344, 1988), and Xen-1,2 (Ruiz i Altaba, *Development* 115: 67, 1992). The embryos were subjected to immunochemistry as described by Turner and Weintraub (*Genes Dev.* 8: 1434, 1994, which is incorporated by reference herein) using primary antibodies detected with alkaline phosphatase-conjugated goat anti-mouse or anti-rabbit antibodies diluted to 1:2000 (Boehringer-Mannheim). Anti-acetylated alpha-tubulin was diluted 1:2000. Anti-Xen-1 was diluted 1:1. Anti-NF-M was diluted 1:2000. Embryos stained for NF-M were fixed in Dent's fixative (20% dimethylsulfoxide/80% methanol) and cleared in 2:1 benzyl benzoate/benzyl alcohol as described by Dent et al. (*Development* 105:61, 1989, which is incorporated by reference herein). *In situ* hybridization of embryos was carried out essentially as described by Harland (in *Methods in Cell Biology*, B.K. Kay, H.J. Pend, Eds, Academic Press, New York, NY, Vol 36, pp. 675-685, 1991, which is incorporated by reference herein) as modified by

Turner and Weintraub (ibid.). *In situ* hybridization with β-tubulin without RNase treatment can also detect tubulin expression in the ciliated epidermal cells. All of these markers displayed ectopic staining on the *neuroD* RNA injected side. Injection of *neuroD* mRNA into vegetal cells led to no ectopic expression of neural markers except in one embryo that showed internal N-CAM staining in the trunk region, suggesting the absence of cofactors or the presence of inhibitors in vegetal cells. However, the one embryo that showed ectopic neurons in the internal organ tissue suggests that it may be possible to convert non-ectodermal lineage cells into neurons under certain conditions.

The embryos were also stained with markers that detect Rohon-Beard cells (cells in which *neuroD* is normally expressed). Immunostaining using the method described above for Rohon-Beard cell-specific markers such as HNK-1 (Nordlander, *Dev. Brain Res.* 50: 147, 1989, which is incorporated by reference herein) at a dilution of 1:1, Islet-1 (Ericson et al., *Science* 256: 1555, 1992 and Korzh et al., *Development* 118: 417, 1993) at a dilution of 1:500, and *in situ* hybridization as described above with shaker-1 (Ribera et al., *J. Neurosci.* 13: 4988, 1993) showed more cells staining on the injected side of the embryos.

The combined results support the notion that ectopic expression of NeuroD induced differentiation of neuronal cells from cells that, without *neuroD* microinjection, would have given rise to non-neuronal cells. In summary, these experiments support the notion that ectopic *neuroD* expression can be used to convert a non-neuronal cell (i.e., uncommitted neural crest cells and epidermal epithelial basal stem cells) into a neuron. These findings offer for the first time the potential for gene therapy to induce neuron formation in injured neural tissues.

Interesting morphological abnormalities were observed in the microinjected embryos. In many cases the eye on the microinjected side of the embryo failed to develop. In other embryos, the spinal cord on the microinjected side of the embryo failed to develop properly, and the tissues were strongly immunopositive when stained with anti-N-CAM. In addition, at the mid-neurula stage many microinjected embryos exhibited an increase in cell mass in the cranial region of the embryo from which (in a normal embryo) the neural crest cells and their derivatives (i.e., cranial ganglionic cells) would migrate. The observed cranial bulge exhibited strong immunostaining with antibodies specific for N-CAM. These results were interpreted to mean that morphological changes in the eye, neural crest, and spinal cord resulted from premature neural differentiation which altered the migration of neural and neural crest precursor cells.

NeuroD-injected embryos were also assayed for alteration in the expression of Xtwist, the *Xenopus* homolog of *Drosophila* twist, to determine whether *neuroD* converted non-neuronal components of neural crest cells into the neural lineage. In wild-type embryos, Xtwist is strongly expressed in the non-neuronal population cephalic neural crest cells that give rise to the connective tissue and skeleton of the head. *NeuroD*-injected embryos were completely missing Xtwist expression in the migrating cranial neural crest cells on the injected side. The failure to generate sufficient cranial mesenchymal neural crest precursors in *neuroD*-injected embryos was also observed morphologically, since many of the injected embryos exhibited poor branchial arch development in the head. Furthermore, the increased mass of cells in the cephalic region stained very strongly for N-CAM, β-tubulin, and Xen-1, indicating that these cells were neural in character.

The converse experiment in which frog embryos were injected with Xtwist mRNA showed that ectopic expression of Xtwist significantly decreased *neuroD* expression on the injected side. Thus, two members of the bHLH family, *neuroD* and Xtwist, may compete for defining the identity of different cell types derived from the neural crest. In the *neuroD*-injected embryos, exogenous *neuroD* may induce premigratory neural crest to differentiate into neurons *in situ*, and consequently they fail to migrate to their normal positions.

The effect of introduction of exogenous *neuroD* on the fate of cells that normally express *neuroD*, such as cranial ganglia, eye, otic vesicle, olfactory organs, and primary neurons, and on other CNS cells that normally do not express *neuroD*, was determined by staining for differentiation markers. When the cranial region of the embryo is severely affected by ectopic *neuroD*, the injected side of the embryos displayed either small or no eyes in addition to poorly organized brains, otic vesicles, and olfactory organs. Moreover, as the embryos grew, the spinal cord showed retarded growth, remaining thinner and shorter on the *neuroD*-injected side.

N-CAM staining in the normal embryo at early stages was not uniform throughout the entire neural plate, but rather was more prominent in the medial region of the neural plate. Injected embryos analyzed for N-CAM expression show that the neural plate on the injected side of the early stage embryos was stained more intensely and more laterally. The increase in N-CAM staining was not associated with any lateral expansion of the neural plate as assayed by visual inspection and staining with the epidermal marker EpA. This was in contrast to what has been observed with XASH-3 injection that causes neural plate expansion. These observations suggest

that the first effects of *neuroD* are to cause neuronal precursors in the neural plate to differentiate prematurely.

To determine whether *neuroD* caused neuronal precursors to differentiate prematurely, injected embryos were stained using two neuronal markers that are expressed in differentiated neurons, neural specific β-tubulin and tanabin. *In situ* hybridization for β-tubulin and tanabin was carried out as described above. Over-expression of *neuroD* dramatically increased the β-tubulin signals in the region of the neural plate containing both motor neurons and Rohon-Beard cells at stage 14. The earliest ectopic β-tubulin positive cells on the injected side were observed at the end of gastrulation when the control side did not yet show any β-tubulin positive cells. Tanabin was also expressed in more cells in the spinal cord in the *neuroD* injected side of the embryos at stage 14. These results suggest that neuroD can cause premature differentiation of the neural precursors into differentiated neurons. This is a powerful indication that, when ectopically expressed or over-expressed, NeuroD can differentiate mitotic cells into non-dividing mature neurons.

EXAMPLE 11

Human genomic *NeuroD.*

Genomic clones encoding human NeuroD were obtained by probing a human fibroblast genomic library with the mouse neuroD cDNA. Host *E. coli* strain LE392 (New England Biolabs) were grown in LB + 10 mM $MgSO_4$ 0.2% maltose overnight at 37°C. The cells were harvested and resuspended in 10 mM $MgSO_4$ to a final OD600 of 2. The resuspended cells were used as hosts for phage infection. The optimal volume of phage stock for use in this screening was determined by using serial dilutions of the phage stock of a human fibroblast genomic library in lambda FIX II (Stratagene) to infect LE392 cells (New England Biolabs). To obtain approximately 50,000 plaques per plate, a 2.5 µl aliquot of the phage stock was used to infect 600 µl of the resuspended LE392 cells. The cells were incubated with the phage for 15 minutes at 37°C, after which the cells were mixed with 6.5 ml of top agar warmed to 50°C. The top agar was plated on solid LB, and incubated overnight at 37°C. A total of 22 15-cm plates were prepared in this manner.

Duplicate plaque lifts were prepared. A first set of Hybond membranes (Amersham) were placed onto the plates and allowed to sit for 2 minutes. The initial membranes were removed and the duplicate membranes were laid on the plates for 4 minutes. The membranes were allowed to air dry; then the phage were denatured in 0.5 M NaOH, 1.5 M NaCl for 7 minutes. The membranes were neutralized with two washes in neutralization buffer (1.5M NaCl, 0.5 M Tris, pH 7.2). After

neutralization, the membranes were crosslinked by exposure to UV. A 1 kb Eco RI-Hind III fragment containing murine *neuroD* coding sequences was random primed using the Random Priming Kit (Boehringer Mannheim) according to the manufacturer's instructions. Membranes were prepared for hybridization by placing six membranes in 10 ml of FBI hybridization buffer [100 g polyethylene glycol 800, 350 ml 20% SDS, 75 ml 20X SSPE; add water to a final volume of one liter.] and incubating the membranes at 65°C for 10 minutes. After 10 minutes, denatured salmon sperm DNA was added to a final concentration of 10 μg/ml and denatured probe was added to a final concentration of 0.25-0.5 x 10^7 cpm/ml. The membranes were hybridized at 65°C for a period of 8 hours to overnight. After incubation, the excess probe was removed, and the membranes were washed first in 2 X SSC, 0.1% SDS for 30 minutes at 50°C. The first wash was followed by a final wash in 0.1 X SSC, 0.1% SDS for 30 minutes at 55°C. Autoradiographs of the membranes were prepared. The first screen identified 55 putative positive plaques. Thirty-one of the plaques were subjected to a secondary screen using the method essentially set forth above. Ten positive clones were identified and subjected to a tertiary screen as described above. Eight positive clones were identified after the tertiary screen.

Phage DNA was prepared from clones 14B1, 9F1, and 20A1. The 14B1 and 20A1 phage DNA were digested with Pst I to isolate the 1.2 kb and 1.6 kb fragments, respectively, that hybridized to the mouse *neuroD* probe. The 9F1 phage DNA was digested with Eco RI and SacI to obtain an approximately 2.2 kb fragment that hybridizes with the mouse *neuroD* probe. The fragments were each subcloned into plasmid Bluescript SK (Stratagene) that had been linearized with the appropriate restriction enzyme(s). The fragments were sequenced using Sequenase Version 2.0 from USB (US Biochemical) and the following primers: the universal primer M13-21, the T7 primer, and the T3 primer. Sequence analysis of clones 9F1, (SEQ ID NO:8) and 14B1 (SEQ ID NO:10) showed a high similarity between the mouse and human coding sequences at both the amino acid and nucleotide level. In addition, while clones 9F1 and 14B1 shared 100% identity in the HLH region at the amino acid level (i.e., residues 117-156 in SEQ ID NO:9 and residues 91-130 in SEQ ID NO:11), they diverged in the amino-terminal of the bHLH. This finding strongly suggests that 14B1 is a member of the NeuroD family of genes. Sequence analysis demonstrates that clone 9F1 has a high degree of homology throughout the sequence region that spans the translation start site to the end of the bHLH region. The 9F1 clone has 100% identity to mouse NeuroD in the HLH region (i.e., residues 117-156 in SEQ ID NO:9 and residues 117-156 in SEQ ID NO:2), and an overall identity of

94%. The 14B1 clone also has 100% identity to the HLH region (i.e., residues 91-130 in SEQ ID NO:11 and residues 117-156 in SEQ ID NO:2), but only 40% identity to 9F1 and 39% identity to mouse NeuroD in the amino-terminal region. This demonstrates that 9F1 is the human homolog of mouse *neuroD*, whereas the strong conservation of the neuroD HLH identifies 14B1 as another member of the *neuroD* HLH subfamily.

EXAMPLE 12

Chromosome mapping of human *neuroD* clones.

FISH karyotyping was performed on fixed metaphase spreads of the microcell hybrids essentially as described (Trask et al., *Am. J. Hum. Genet.* 48: 1-15, 1991; and Brandriff et al., *Genomics* 10: 75-82, 1991, which are incorporated by reference herein in their entirety). *NeuroD* sequences were detected using the 9F1 or 20A1 phage DNA as probes labeled using digoxigenin dUTP (Boehringer Mannheim) according to the manufacturer's instructions. Phage DNA was biotinylated by random priming (Gibco/BRL BioNick Kit) and hybridized *in situ* to denatured metaphase chromosome spreads for 24-48 hours. Probes were detected with rhodamine-conjugated antibodies to digoxigenin and chromosomes were counterstained DAPI (Sigma). Signals were viewed through a fluorescence microscope and photographs were taken with color slide film. FISH analysis indicated clone 9F1 maps to human chromosome 2q, and clone 20A1 maps to human chromosome 5.

Chromosome mapping was also carried out on a human/rodent somatic cell hybrid panel (National Institute of General Medical Sciences; Camden NJ). This panel consists of DNA isolated from 24 human/rodent somatic cell hybrids retaining one human chromosomes. For one set of experiments, the panel of DNA's were digested with Eco RI and electrophoresed on an agarose gel. The DNA was transferred to Hybond-N membranes (Amersham). A random primed (Boehringer Mannheim) 4 kb Eco RI-Sac I fragment of clone 9F1 was prepared. The filter was prehybridized in 10 ml of FBI hybridization buffer (see above) at 65°C for 10 minutes. After prehybridization, denatured salmon sperm DNA was added to a final concentration of 10 µg/ml; denatured probe was added to a final concentration of one million cpm/ml. The filter was hybridized at 65°C for a period of 8 hours to overnight. After incubation, excess probe was removed, and the filter was washed first in 2 X SSC, 0.1% SDS for 30 minutes at 65°C. The first wash was followed by a final wash in 0.1 X SSC, 0.1% SDS for 30 minutes at 65°C. An autoradiograph of the filter was prepared. Autoradiographs confirmed the FISH mapping results.

In the second experiment, the panel was digested with Pst I, electrophoresed and transferred essentially as described above. A random-primed (Boehringer Mannheim) 1.6 kb Pst I fragment of clone 20A1 was prepared. The membrane was prehybridized, hybridized with the 20A1 probe and washed as described above. Autoradiographs of the Southern showed that 20A1 mapped to human chromosome 5 and confirmed the FISH mapping results. After autoradiography, the 20A1-probed membrane was stripped by a wash in 0.5 M NaOH, 1.5 M NaCl. The membrane was neutralized in 0.5 M Tris-HCl (pH 7.4), 1.5 M NaCl. The filter was washed in 0.1 X SSC before prehybridization. A random-primed (Boehringer Mannheim) 1.2 kb Pst I fragment of clone 14B1 was prepared. The washed membrane was prehybridized and hybridized with the 14B1 probe as described above. After washing under the previously described conditions, the membrane was autoradiographed. Autoradiographs demonstrated that clone 14B1 mapped to chromosome 17.

EXAMPLE 13

Human *neuroD* complementary DNA.

To obtain a human *neuroD* cDNA, one million plaque forming units (pfu) were plated onto twenty LB + 10 mM $MgSO_4$ (150 mm) plates using the bacterial strain XL-1 Blue (Stratagene). Plating and membrane lifts were performed using standard methods, as described in Example 11. After UV cross-linking, the membranes were pre-hybridized in an aqueous hybridization solution (1% bovine serum albumin, 1 mM EDTA, 0.5 M Na_2HPO_4 (pH 7.4), 7% SDS) at 50°C for two hours.

The *neuroD* cDNA insert was prepared by digesting the pKS+ m7a RX plasmid with Eco RI and Xho I, and isolating the fragment containing the cDNA by electroelution. A probe was made with the cDNA containing fragment by random primed synthesis with random hexanucleotides, dGTP,dATP, dTTP, alpha-^{32}P-labeled dCTP, and Klenow in a buffered solution (25 mM Tris (pH6.9), 50mM KCl, 5mM $MgCl_2$, 1mM DTT). The probe was purified from the unincorporated nucleotides on a G-50 sepharose column. The purified probe was heat denatured at 90°C for 3 minutes.

After prehybridization, the denatured probe was added to the membranes in hybridization solution. The membranes were hybridized for 24 hours at 50°C. Excess probe was removed from the membranes, and the membranes were washed in 0.1 X SSC, 0.1% SDS for 20 minutes at 50°C. The wash solution was changed five times. The membranes were blotted dry and covered with plastic film before being subjected

to autoradiography. Autoradiography of the filters identified 68 positive clones. The clones are plaque-purified and rescreened to obtain pure, positive clones.

The plasmid vector containing cDNA insert was excised *in vivo* from the lambda phage clone according to the Strategene methodology. Briefly, eluted phage and XL-1 Blue cells (200 microliters of OD 600=1) were mixed with R408 helper phage provided by Strategene for 15 minutes at 37°C. Five milliliters of rich bacterial growth media (2 X YT, see Sambrook et al., ibid.) was added, and the cultures were incubated for 3 hours at 37°C. The tubes were heated at 70°C for 20 minutes and spun for 5 minutes at 4,000 X g. After centrifugation, 200 microliters of supernant was added to the same volume of XL-1 Blue cells (OD=1), and the mixture was incubated for 15 minutes at 37°C, after which the bacterial cells were plated onto LB plates containing 50 μg/ml ampicillin. Each colony was picked and grown for sequencing template preparation. The clones are sequenced and compared to the human genomic sequence.

From the foregoing it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modification may be made without deviating from the spirit and scope of the invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Weintraub, Harold
Lee, Jacqueline E.
Tapscott, Stephen J.
Hollenberg, Stanley M.

(ii) TITLE OF INVENTION: Neurogenic Differentiation (NeuroD) Gene
and Protein

(iii) NUMBER OF SEQUENCES: 11

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Christensen O'Connor Johnson Kindness
(B) STREET: 1420 Fifth Avenue, Suite 2800
(C) CITY: Seattle
(D) STATE: WA
(E) COUNTRY: USA
(F) ZIP: 98101-2347

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Broderick, Thomas F.
(B) REGISTRATION NUMBER: 31,332
(C) REFERENCE/DOCKET NUMBER: FHCR-1-8504

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 206-682-8100
(B) TELEFAX: 206-225-0709

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2089 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 229..1302

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ACTACGCAGC ACCGAGGTAC AGACACGCCA GCATGAAGCA CTGCGTTTAA CTTTTCCTGG
60

AGGCATCCAT TTTGCAGTGG ACTCCTGTGT ATTTCTATTT GTGTGCATTT CTGTAGGATT
120

AGGGAGAGGG AGCTGAAGGC TTATCCAGCT TTTAAATATA GCGGGTGGAT TTCCCCCCCT
180

TTCTTCTTCT GCTTGCCTCT CTCCCTGTTC AATACAGGAA GTGGAAAC ATG ACC AAA
237
Met Thr Lys
1

TCA TAC AGC GAG AGC GGG CTG ATG GGC GAG CCT CAG CCC CAA GGT CCC
285
Ser Tyr Ser Glu Ser Gly Leu Met Gly Glu Pro Gln Pro Gln Gly Pro
5 10 15

CCA AGC TGG ACA GAT GAG TGT CTC AGT TCT CAG GAC GAG GAA CAC GAG
333
Pro Ser Trp Thr Asp Glu Cys Leu Ser Ser Gln Asp Glu Glu His Glu
20 25 30 35

GCA GAC AAG AAA GAG GAC GAG CTT GAA GCC ATG AAT GCA GAG GAG GAC
381
Ala Asp Lys Lys Glu Asp Glu Leu Glu Ala Met Asn Ala Glu Glu Asp
40 45 50

TCT CTG AGA AAC GGG GGA GAG GAG GAG GAG GAA GAT GAG GAT CTA GAG
429
Ser Leu Arg Asn Gly Gly Glu Glu Glu Glu Glu Asp Glu Asp Leu Glu
55 60 65

GAA GAG GAG GAA GAA GAA GAG GAG GAG GAG GAT CAA AAG CCC AAG AGA
477
Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Asp Gln Lys Pro Lys Arg
70 75 80

CGG GGT CCC AAA AAG AAA AAG ATG ACC AAG GCG CGC CTA GAA CGT TTT
525
Arg Gly Pr Lys Lys Lys Lys Met Thr Lys Ala Arg Leu Glu Arg Phe
85 90 95

```
AAA TTA AGG CGC ATG AAG GCC AAC GCC CGC GAG CGG AAC CGC ATG CAC
573
Lys Leu Arg Arg Met Lys Ala Asn Ala Arg Glu Arg Asn Arg Met His
100                 105                 110                 115

GGG CTG AAC GCG GCG CTG GAC AAC CTG CGC AAG GTG GTA CCT TGC TAC
621
Gly Leu Asn Ala Ala Leu Asp Asn Leu Arg Lys Val Val Pro Cys Tyr
                120                 125                 130

TCC AAG ACC CAG AAA CTG TCT AAA ATA GAG ACA CTG CGC TTG GCC AAG
669
Ser Lys Thr Gln Lys Leu Ser Lys Ile Glu Thr Leu Arg Leu Ala Lys
            135                 140                 145

AAC TAC ATC TGG GCT CTG TCA GAG ATC CTG CGC TCA GGC AAA AGC CCT
717
Asn Tyr Ile Trp Ala Leu Ser Glu Ile Leu Arg Ser Gly Lys Ser Pro
        150                 155                 160

GAT CTG GTC TCC TTC GTA CAG ACG CTC TGC AAA GGT TTG TCC CAG CCC
765
Asp Leu Val Ser Phe Val Gln Thr Leu Cys Lys Gly Leu Ser Gln Pro
    165                 170                 175

ACT ACC AAT TTG GTC GCC GGC TGC CTG CAG CTC AAC CCT CGG ACT TTC
813
Thr Thr Asn Leu Val Ala Gly Cys Leu Gln Leu Asn Pro Arg Thr Phe
180                 185                 190                 195

TTG CCT GAG CAG AAC CCG GAC ATG CCC CCG CAT CTG CCA ACC GCC AGC
861
Leu Pro Glu Gln Asn Pro Asp Met Pro Pro His Leu Pro Thr Ala Ser
                200                 205                 210

GCT TCC TTC CCG GTG CAT CCC TAC TCC TAC CAG TCC CCT GGA CTG CCC
909
Ala Ser Phe Pro Val His Pro Tyr Ser Tyr Gln Ser Pro Gly Leu Pro
            215                 220                 225

AGC CCG CCC TAC GGC ACC ATG GAC AGC TCC CAC GTC TTC CAC GTC AAG
957
Ser Pro Pro Tyr Gly Thr Met Asp Ser Ser His Val Phe His Val Lys
        230                 235                 240

CCG CCG CCA CAC GCC TAC AGC GCA GCT CTG GAG CCC TTC TTT GAA AGC
1005
Pro Pro Pro His Ala Tyr Ser Ala Ala Leu Glu Pro Phe Phe Glu Ser
    245                 250                 255

CCC CTA ACT GAC TGC ACC AGC CCT TCC TTT GAC GGA CCC CTC AGC CCG
1053
Pr  Leu Thr Asp Cys Thr Ser Pro Ser Phe Asp Gly Pro Leu Ser Pro
260                 265                 270                 275
```

```
CCG CTC AGC ATC AAT GGC AAC TTC TCT TTC AAA CAC GAA CCA TCC GCC
1101
Pro Leu Ser Ile Asn Gly Asn Phe Ser Phe Lys His Glu Pro Ser Ala
                280                 285                 290

GAG TTT GAA AAA AAT TAT GCC TTT ACC ATG CAC TAC CCT GCA GCG ACG
1149
Glu Phe Glu Lys Asn Tyr Ala Phe Thr Met His Tyr Pro Ala Ala Thr
            295                 300                 305

CTG GCA GGG CCC CAA AGC CAC GGA TCA ATC TTC TCT TCC GGT GCC GCT
1197
Leu Ala Gly Pro Gln Ser His Gly Ser Ile Phe Ser Ser Gly Ala Ala
        310                 315                 320

GCC CCT CGC TGC GAG ATC CCC ATA GAC AAC ATT ATG TCT TTC GAT AGC
1245
Ala Pro Arg Cys Glu Ile Pro Ile Asp Asn Ile Met Ser Phe Asp Ser
    325                 330                 335

CAT TCG CAT CAT GAG CGA GTC ATG AGT GCC CAG CTT AAT GCC ATC TTT
1293
His Ser His His Glu Arg Val Met Ser Ala Gln Leu Asn Ala Ile Phe
340                 345                 350                 355

CAC GAT TAGAGGGCAC GTCAGTTTCA CTATTCCCGG GAAACGAATC CACTGTGCGT
1349
His Asp

ACAGTGACTG TCCTGTTTAC AGAAGGCAGC CCTTTTGCTA AGATTGCTGC AAAGTGCAAA
1409

TACTCAAAGC TTCAAGTGAT ATATGTATTT ATTGTCGTTA CTGCCTTTGG AAGAAACAGG
1469

GGATCAAAGT TCCTGTTCAC CTTATGTATT GTTTTCTATA GCTCTTCTAT TTTAAAAATA
1529

ATAATACAGT AAAGTAAAAA AGAAAATGTG TACCACGAAT TTCGTGTAGC TGTATTCAGA
1589

TCGTATTAAT TATCTGATCG GGATAAAAAA AATCACAAGC AATAATTAGG ATCTATGCAA
1649

TTTTTAAACT AGTAATGGGC CAATTAAAAT ATATATAAAT ATATATTTTT CAACCAGCAT
1709

TTTACTACCT GTGACCTTTC CCATGCTGAA TTATTTTGTT GTGATTTTGT ACAGAATTTT
1769

TAATGACTTT TTATAACGTG GATTTCCTAT TTTAAAACCA TGCAGCTTCA TCAATTTTTA
1829
```

TACATATCAG AAAAGTAGAA TTATATCTAA TTTATACAAA ATAATTTAAC TAATTTAAAC
1889

CAGCAGAAAA GTGCTTAGAA AGTTATTGCG TTGCCTTAGC ACTTCTTTCT TCTCTAATTG
1949

TAAAAAAGAA AAAAAAAAAA AAAAAACTCG AGGGGGGGCC CGGTACCCAG CTTTTGTTCC
2009

CTTTAGTGAG GGTTAATTGC GCGCTTGGCG TAATCATGGT CATAGCTGTT TCCTGTGTGA
2069

ATTGTTATCC GCTCACAATT
2089

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 357 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Thr Lys Ser Tyr Ser Glu Ser Gly Leu Met Gly Glu Pro Gln Pro
1 5 10 15

Gln Gly Pro Pro Ser Trp Thr Asp Glu Cys Leu Ser Ser Gln Asp Glu
20 25 30

Glu His Glu Ala Asp Lys Lys Glu Asp Glu Leu Glu Ala Met Asn Ala
35 40 45

Glu Glu Asp Ser Leu Arg Asn Gly Gly Glu Glu Glu Glu Asp Glu
50 55 60

Asp Leu Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Asp Gln Lys
65 70 75 80

Pro Lys Arg Arg Gly Pro Lys Lys Lys Lys Met Thr Lys Ala Arg Leu
85 90 95

Glu Arg Phe Lys Leu Arg Arg Met Lys Ala Asn Ala Arg Glu Arg Asn
100 105 110

Arg Met His Gly Leu Asn Ala Ala Leu Asp Asn Leu Arg Lys Val Val
115 120 125

Pro Cys Tyr Ser Lys Thr Gln Lys Leu Ser Lys Ile Glu Thr Leu Arg
130 135 140

```
Leu Ala Lys Asn Tyr Ile Trp Ala Leu Ser Glu Ile Leu Arg Ser Gly
145                 150                 155                 160

Lys Ser Pro Asp Leu Val Ser Phe Val Gln Thr Leu Cys Lys Gly Leu
                165                 170                 175

Ser Gln Pro Thr Thr Asn Leu Val Ala Gly Cys Leu Gln Leu Asn Pro
            180                 185                 190

Arg Thr Phe Leu Pro Glu Gln Asn Pro Asp Met Pro Pro His Leu Pro
        195                 200                 205

Thr Ala Ser Ala Ser Phe Pro Val His Pro Tyr Ser Tyr Gln Ser Pro
    210                 215                 220

Gly Leu Pro Ser Pro Pro Tyr Gly Thr Met Asp Ser Ser His Val Phe
225                 230                 235                 240

His Val Lys Pro Pro Pro His Ala Tyr Ser Ala Ala Leu Glu Pro Phe
                245                 250                 255

Phe Glu Ser Pro Leu Thr Asp Cys Thr Ser Pro Ser Phe Asp Gly Pro
            260                 265                 270

Leu Ser Pro Pro Leu Ser Ile Asn Gly Asn Phe Ser Phe Lys His Glu
        275                 280                 285

Pro Ser Ala Glu Phe Glu Lys Asn Tyr Ala Phe Thr Met His Tyr Pro
    290                 295                 300

Ala Ala Thr Leu Ala Gly Pro Gln Ser His Gly Ser Ile Phe Ser Ser
305                 310                 315                 320

Gly Ala Ala Ala Pro Arg Cys Glu Ile Pro Ile Asp Asn Ile Met Ser
                325                 330                 335

Phe Asp Ser His Ser His His Glu Arg Val Met Ser Ala Gln Leu Asn
            340                 345                 350

Ala Ile Phe His Asp
        355
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1275 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Xenopus laevis

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 25..1083

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
ATTTCCTTTC TCCAGATCTA AAAA ATG ACC AAA TCG TAT GGA GAG AAT GGG
51
                           Met Thr Lys Ser Tyr Gly Glu Asn Gly
                             1               5

CTG ATC CTG GCC GAG ACT CCG GGC TGC AGA GGA TGG GTG GAC GAA TGC
99
Leu Ile Leu Ala Glu Thr Pro Gly Cys Arg Gly Trp Val Asp Glu Cys
 10                  15                  20                  25

CTG AGT TCT CAG GAT GAA AAC GAT CTG GAG AAA AAG GAG GGA GAG TTG
147
Leu Ser Ser Gln Asp Glu Asn Asp Leu Glu Lys Lys Glu Gly Glu Leu
                 30                  35                  40

ATG AAA GAA GAC GAT GAA GAC TCA CTG AAT CAT CAC AAT GGA GAG GAG
195
Met Lys Glu Asp Asp Glu Asp Ser Leu Asn His His Asn Gly Glu Glu
             45                  50                  55

AAC GAG GAA GAG GAT GAA GGG GAT GAG GAG GAG GAG GAC GAT GAA GAT
243
Asn Glu Glu Glu Asp Glu Gly Asp Glu Glu Glu Glu Asp Asp Glu Asp
         60                  65                  70

GAT GAT GAG GAT GAC GAC CAG AAA CCC AAA AGG CGA GGA CCG AAA AAG
291
Asp Asp Glu Asp Asp Asp Gln Lys Pro Lys Arg Arg Gly Pro Lys Lys
     75                  80                  85

AAA AAA ATG ACG AAA GCC CGG GTG GAG CGA TTT AAA GTG AGA CGC ATG
339
Lys Lys Met Thr Lys Ala Arg Val Glu Arg Phe Lys Val Arg Arg Met
 90                  95                  100                 105

AAG GCA AAC GCC AGG GAG AGG AAT CGC ATG CAC GGA CTC AAC GAT GCC
387
Lys Ala Asn Ala Arg Glu Arg Asn Arg Met His Gly Leu Asn Asp Ala
                110                 115                 120

CTG GAC AGT CTG CGC AAA GTT GTG CCC TGC TAC TCC AAA ACA CAA AAG
435
Leu Asp Ser Leu Arg Lys Val Val Pro Cys Tyr Ser Lys Thr Gln Lys
            125                 130                 135
```

```
TTG TCT AAG ATT GAA ACT CTG CGC CTG GCT AAG AAC TAC ATC TGG GCT
483
Leu Ser Lys Ile Glu Thr Leu Arg Leu Ala Lys Asn Tyr Ile Trp Ala
        140                 145                 150

CTT TCT GAG ATT TTA AGG TCC GGC AAA AGC CCA GAC CTG GTG TCC TTT
531
Leu Ser Glu Ile Leu Arg Ser Gly Lys Ser Pro Asp Leu Val Ser Phe
    155                 160                 165

GTA CAA ACT CTC TGC AAA GGT TTG TCG CAG CCC ACC ACC AAT CTA GTA
579
Val Gln Thr Leu Cys Lys Gly Leu Ser Gln Pro Thr Thr Asn Leu Val
170                 175                 180                 185

GCG GGG TGT CTG CAG CTG AAC CCC AGA ACT TTC CTT CCT GAG CAG AGT
627
Ala Gly Cys Leu Gln Leu Asn Pro Arg Thr Phe Leu Pro Glu Gln Ser
                190                 195                 200

CAG GAC ATC CAG TCG CAC ATG CAA ACA GCG AGC TCT TCC TTC CCT CTG
675
Gln Asp Ile Gln Ser His Met Gln Thr Ala Ser Ser Ser Phe Pro Leu
            205                 210                 215

CAG GGC TAT CCC TAT CAG TCC CCT GGT CTT CCC AGT CCC CCC TAT GGT
723
Gln Gly Tyr Pro Tyr Gln Ser Pro Gly Leu Pro Ser Pro Pro Tyr Gly
        220                 225                 230

ACC ATG GAC AGC TCC CAT GTA TTC CAC GTC AAG CCT CAC TCC TAT GGG
771
Thr Met Asp Ser Ser His Val Phe His Val Lys Pro His Ser Tyr Gly
    235                 240                 245

GCG GCC CTG GAG CCT TTC TTT GAC AGC AGC ACC GTC ACT GAG TGT ACC
819
Ala Ala Leu Glu Pro Phe Phe Asp Ser Ser Thr Val Thr Glu Cys Thr
250                 255                 260                 265

AGC CCG TCA TTC GAT GGT CCC CTG AGC CCA CCC CTT AGT GTT AAT GGG
867
Ser Pro Ser Phe Asp Gly Pro Leu Ser Pro Pro Leu Ser Val Asn Gly
                270                 275                 280

AAC TTT ACT TTT AAA CAC GAG CAT TCG GAG TAT GAT AAA AAT TAC ACG
915
Asn Phe Thr Phe Lys His Glu His Ser Glu Tyr Asp Lys Asn Tyr Thr
            285                 290                 295

TTC ACT ATG CAC TAT CCT GCA GCC ACT ATA TCC CAG GGC CAC GGA CCA
963
Phe Thr Met His Tyr Pro Ala Ala Thr Ile Ser Gln Gly His Gly Pro
```

```
        300                 305                 310

TTG TTC TCC ACG GGG GGA CCA CGC TGT GAA ATC CCA ATA GAC ACC ATC
1011
Leu Phe Ser Thr Gly Gly Pro Arg Cys Glu Ile Pro Ile Asp Thr Ile
    315                 320                 325

ATG TCC TAT GAC GGT CAC TCC CAC CAT GAA AGA GTC ATG AGT GCC CAG
1059
Met Ser Tyr Asp Gly His Ser His His Glu Arg Val Met Ser Ala Gln
330                 335                 340                 345

CTA AAT GCC ATC TTT CAT GAT TAACCCTTGG AAGATCAAAA CAACTGACTG
1110
Leu Asn Ala Ile Phe His Asp
                350

TGCATTGCCA GGACTGTCTT GTTTACCAAG GGCAGACACG TGGGTAGTAA AAGTGCAAAT
1170

GCCCCACTCT GGGGCTGTAA CAAACTTGAT CTTGTCCTGC CTTTAGATAT GGGGAAACCT
1230

AATGTATTAA TTCCCACCTC CTTCCAATCG ACACTCCTTT AAATT
1275
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 352 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Thr Lys Ser Tyr Gly Glu Asn Gly Leu Ile Leu Ala Glu Thr Pro
  1               5                  10                  15

Gly Cys Arg Gly Trp Val Asp Glu Cys Leu Ser Ser Gln Asp Glu Asn
             20                  25                  30

Asp Leu Glu Lys Lys Glu Gly Glu Leu Met Lys Glu Asp Asp Glu Asp
         35                  40                  45

Ser Leu Asn His His Asn Gly Glu Glu Asn Glu Glu Glu Asp Glu Gly
     50                  55                  60

Asp Glu Glu Glu Glu Asp Asp Glu Asp Asp Asp Glu Asp Asp Asp Gln
 65                  70                  75                  80

Lys Pro Lys Arg Arg Gly Pro Lys Lys Lys Lys Met Thr Lys Ala Arg
```

85 90 95

Val Glu Arg Phe Lys Val Arg Arg Met Lys Ala Asn Ala Arg Glu Arg
100 105 110

Asn Arg Met His Gly Leu Asn Asp Ala Leu Asp Ser Leu Arg Lys Val
115 120 125

Val Pro Cys Tyr Ser Lys Thr Gln Lys Leu Ser Lys Ile Glu Thr Leu
130 135 140

Arg Leu Ala Lys Asn Tyr Ile Trp Ala Leu Ser Glu Ile Leu Arg Ser
145 150 155 160

Gly Lys Ser Pro Asp Leu Val Ser Phe Val Gln Thr Leu Cys Lys Gly
165 170 175

Leu Ser Gln Pro Thr Thr Asn Leu Val Ala Gly Cys Leu Gln Leu Asn
180 185 190

Pro Arg Thr Phe Leu Pro Glu Gln Ser Gln Asp Ile Gln Ser His Met
195 200 205

Gln Thr Ala Ser Ser Ser Phe Pro Leu Gln Gly Tyr Pro Tyr Gln Ser
210 215 220

Pro Gly Leu Pro Ser Pro Pro Tyr Gly Thr Met Asp Ser Ser His Val
225 230 235 240

Phe His Val Lys Pro His Ser Tyr Gly Ala Ala Leu Glu Pro Phe Phe
245 250 255

Asp Ser Ser Thr Val Thr Glu Cys Thr Ser Pro Ser Phe Asp Gly Pro
260 265 270

Leu Ser Pro Pro Leu Ser Val Asn Gly Asn Phe Thr Phe Lys His Glu
275 280 285

His Ser Glu Tyr Asp Lys Asn Tyr Thr Phe Thr Met His Tyr Pro Ala
290 295 300

Ala Thr Ile Ser Gln Gly His Gly Pro Leu Phe Ser Thr Gly Gly Pro
305 310 315 320

Arg Cys Glu Ile Pro Ile Asp Thr Ile Met Ser Tyr Asp Gly His Ser
325 330 335

His His Glu Arg Val Met Ser Ala Gln Leu Asn Ala Ile Phe His Asp
340 345 350

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Asn Ala Arg Glu Arg Arg Arg
1 5

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Asn Glu Arg Glu Arg Asn Arg
1 5

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Asn Ala Arg Glu Arg
1 5

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 524 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(vii) IMMEDIATE SOURCE:
(B) CLONE: 9F1

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 57..524

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
TTTTTCTGCT TTTCTTTCTG TTTGCCTCTC CCTTGTTGAA TGTAGGAAAT CGAAAC
56

ATG ACC AAA TCG TAC AGC GAG AGT GGG CTG ATG GGC GAG CCT CAG CCC
104
Met Thr Lys Ser Tyr Ser Glu Ser Gly Leu Met Gly Glu Pro Gln Pro
  1               5                  10                  15

CAA GGT CCT CCA AGC TGG ACA GAC GAG TGT CTC AGT TCT CAG GAC GAG
152
Gln Gly Pro Pro Ser Trp Thr Asp Glu Cys Leu Ser Ser Gln Asp Glu
            20                  25                  30

GAG CAC GAG GCA GAC AAG AAG GAG GAC GAC CTC GAA GCC ATG AAC GCA
200
Glu His Glu Ala Asp Lys Lys Glu Asp Asp Leu Glu Ala Met Asn Ala
        35                  40                  45

GAG GAG GAC TCA CTG AGG AAC GGG GGA GAG GAG GAG GAC GAA GAT GAG
248
Glu Glu Asp Ser Leu Arg Asn Gly Gly Glu Glu Glu Asp Glu Asp Glu
    50                  55                  60

GAC CTG GAA GAG GAG GAA GAA GAG GAA GAG GAG GAT GAC GAT CAA AAG
296
Asp Leu Glu Glu Glu Glu Glu Glu Glu Glu Glu Asp Asp Asp Gln Lys
 65                  70                  75                  80

CCC AAG AGA CGC GGC CCC AAA AAG AAG AAG ATG ACT AAG GCT CGC CTG
344
Pro Lys Arg Arg Gly Pro Lys Lys Lys Lys Met Thr Lys Ala Arg Leu
```

```
                85                  90                  95

GAG CGT TTT AAA TTG AGA CGC ATG AAG GCT AAC GCC CGG GAG CGG AAC
392
Glu Arg Phe Lys Leu Arg Arg Met Lys Ala Asn Ala Arg Glu Arg Asn
            100                 105                 110

CGC ATG CAC GGA CTG AAC GCG GCG CTA GAC AAC CTG CGC AAG GTG GTG
440
Arg Met His Gly Leu Asn Ala Ala Leu Asp Asn Leu Arg Lys Val Val
        115                 120                 125

CCT TGC TAT TCT AAG ACG CAG AAG CTG TCC AAA ATC GAG ACT CTG CGC
488
Pro Cys Tyr Ser Lys Thr Gln Lys Leu Ser Lys Ile Glu Thr Leu Arg
    130                 135                 140

TTG GCC AAG AAC TAC ATC TGG GCT CTG TCG GAG ATC
524
Leu Ala Lys Asn Tyr Ile Trp Ala Leu Ser Glu Ile
145                 150                 155
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 156 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
Met Thr Lys Ser Tyr Ser Glu Ser Gly Leu Met Gly Glu Pro Gln Pro
  1               5                  10                  15

Gln Gly Pro Pro Ser Trp Thr Asp Glu Cys Leu Ser Ser Gln Asp Glu
            20                  25                  30

Glu His Glu Ala Asp Lys Lys Glu Asp Asp Leu Glu Ala Met Asn Ala
        35                  40                  45

Glu Glu Asp Ser Leu Arg Asn Gly Gly Glu Glu Glu Asp Glu Asp Glu
    50                  55                  60

Asp Leu Glu Glu Glu Glu Glu Glu Glu Glu Glu Asp Asp Asp Gln Lys
65                  70                  75                  80

Pro Lys Arg Arg Gly Pro Lys Lys Lys Lys Met Thr Lys Ala Arg Leu
                85                  90                  95

Glu Arg Phe Lys Leu Arg Arg Met Lys Ala Asn Ala Arg Glu Arg Asn
            100                 105                 110
```

Arg Met His Gly Leu Asn Ala Ala Leu Asp Asn Leu Arg Lys Val Val
115 120 125

Pro Cys Tyr Ser Lys Thr Gln Lys Leu Ser Lys Ile Glu Thr Leu Arg
130 135 140

Leu Ala Lys Asn Tyr Ile Trp Ala Leu Ser Glu Ile
145 150 155

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 485 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(vii) IMMEDIATE SOURCE:
(B) CLONE: 14B1

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..485

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GG GCC AGG GGC TCC GGG GCC AGC CCG GGC GGC CAA GCC AGT CCC TCT
47
Ala Arg Gly Ser Gly Ala Ser Pro Gly Gly Gln Ala Ser Pro Ser
1 5 10 15

CCG TGG AGA AGA GGG GAC GGA GGC CAC GTT GGC CGA GGT CAA GGA GGA
95
Pro Trp Arg Arg Gly Asp Gly Gly His Val Gly Arg Gly Gln Gly Gly
20 25 30

AGG CGG CTG GGG GGA GAG GAG GAG GAG GAA GAG GAG GAG GAA GAA GGA
143
Arg Arg Leu Gly Gly Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Gly
35 40 45

CTG GAC GAG GCG GAG GGC GAG CGG CCC AAG AAG CGC GGG CCC AAG AAG
191
Leu Asp Glu Ala Glu Gly Glu Arg Pro Lys Lys Arg Gly Pro Lys Lys
50 55 60

```
CGC AAG ATG ACC AAG GCG CGC TTG GAG CGC TCC AAG CTT CGG CGG CAG
239
Arg Lys Met Thr Lys Ala Arg Leu Glu Arg Ser Lys Leu Arg Arg Gln
 65                  70                  75

AAG GCG AAC GCG CGG GAG AAC CGC ATG CAC GAC CTG AAC GCA GCC CTG
287
Lys Ala Asn Ala Arg Glu Asn Arg Met His Asp Leu Asn Ala Ala Leu
 80                  85                  90                  95

GAC AAC CTG CGC AAG GTG GTG CCC TGC TAC TCC AAG ACG CAG AAG CTG
335
Asp Asn Leu Arg Lys Val Val Pro Cys Tyr Ser Lys Thr Gln Lys Leu
                100                 105                 110

TCC AAG ATC GAG ACG CTG CGC CTA GCC AAG AAC TAT ATC TGG GCG CTC
383
Ser Lys Ile Glu Thr Leu Arg Leu Ala Lys Asn Tyr Ile Trp Ala Leu
            115                 120                 125

TCG GAG ATC CTG CGC TCC GGC AAG CGG CCA GAC CTA GTG TCC TAC GTG
431
Ser Glu Ile Leu Arg Ser Gly Lys Arg Pro Asp Leu Val Ser Tyr Val
        130                 135                 140

CAG ACT CTG TGC AAG GGT CTG TCG CAG CCC ACC ACC AAT CTG GTG GCC
479
Gln Thr Leu Cys Lys Gly Leu Ser Gln Pro Thr Thr Asn Leu Val Ala
    145                 150                 155

GGC TGT
485
Gly Cys
160
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 161 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
Ala Arg Gly Ser Gly Ala Ser Pro Gly Gly Gln Ala Ser Pro Ser Pro
 1               5                  10                  15

Trp Arg Arg Gly Asp Gly Gly His Val Gly Arg Gly Gln Gly Gly Arg
            20                  25                  30

Arg Leu Gly Gly Glu Glu Glu Glu Glu Glu Glu Glu Glu Gly Leu
```

35 40 45

Asp Glu Ala Glu Gly Glu Arg Pro Lys Lys Arg Gly Pro Lys Lys Arg
50 55 60

Lys Met Thr Lys Ala Arg Leu Glu Arg Ser Lys Leu Arg Arg Gln Lys
65 70 75 80

Ala Asn Ala Arg Glu Asn Arg Met His Asp Leu Asn Ala Ala Leu Asp
85 90 95

Asn Leu Arg Lys Val Val Pro Cys Tyr Ser Lys Thr Gln Lys Leu Ser
100 105 110

Lys Ile Glu Thr Leu Arg Leu Ala Lys Asn Tyr Ile Trp Ala Leu Ser
115 120 125

Glu Ile Leu Arg Ser Gly Lys Arg Pro Asp Leu Val Ser Tyr Val Gln
130 135 140

Thr Leu Cys Lys Gly Leu Ser Gln Pro Thr Thr Asn Leu Val Ala Gly
145 150 155 160

Cys

The embodiments of the invention in which an exclusive property or privilege is claimed are defined as follows:

1. An isolated nucleic acid molecule which comprises at least 15 nucleotides and which hybridizes under stringent conditions with a *neuroD* HLH domain selected from among nucleotides 577-696 of SEQ ID NO:1, nucleotides 376-495 of SEQ ID NO:3, nucleotides 405-524 of SEQ ID NO:8, nucleotides 273-392 of SEQ ID NO:10, and complements thereof.

2. A vector comprising in serial array a promoter, the nucleic acid molecule of claim 1, and a poly(A) tail.

3. A cell transformed by the nucleic acid molecule of claim 1.

4. An isolated nucleic acid molecule of claim 1, which hybridizes under stringent conditions with a nucleic acid molecule selected from among SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:8, SEQ ID NO:10, and complements thereof.

5. A recombinant peptide encoded by the nucleic acid molecule of claim 1.

6. A recombinant peptide encoded by the nucleic acid molecule of claim 4.

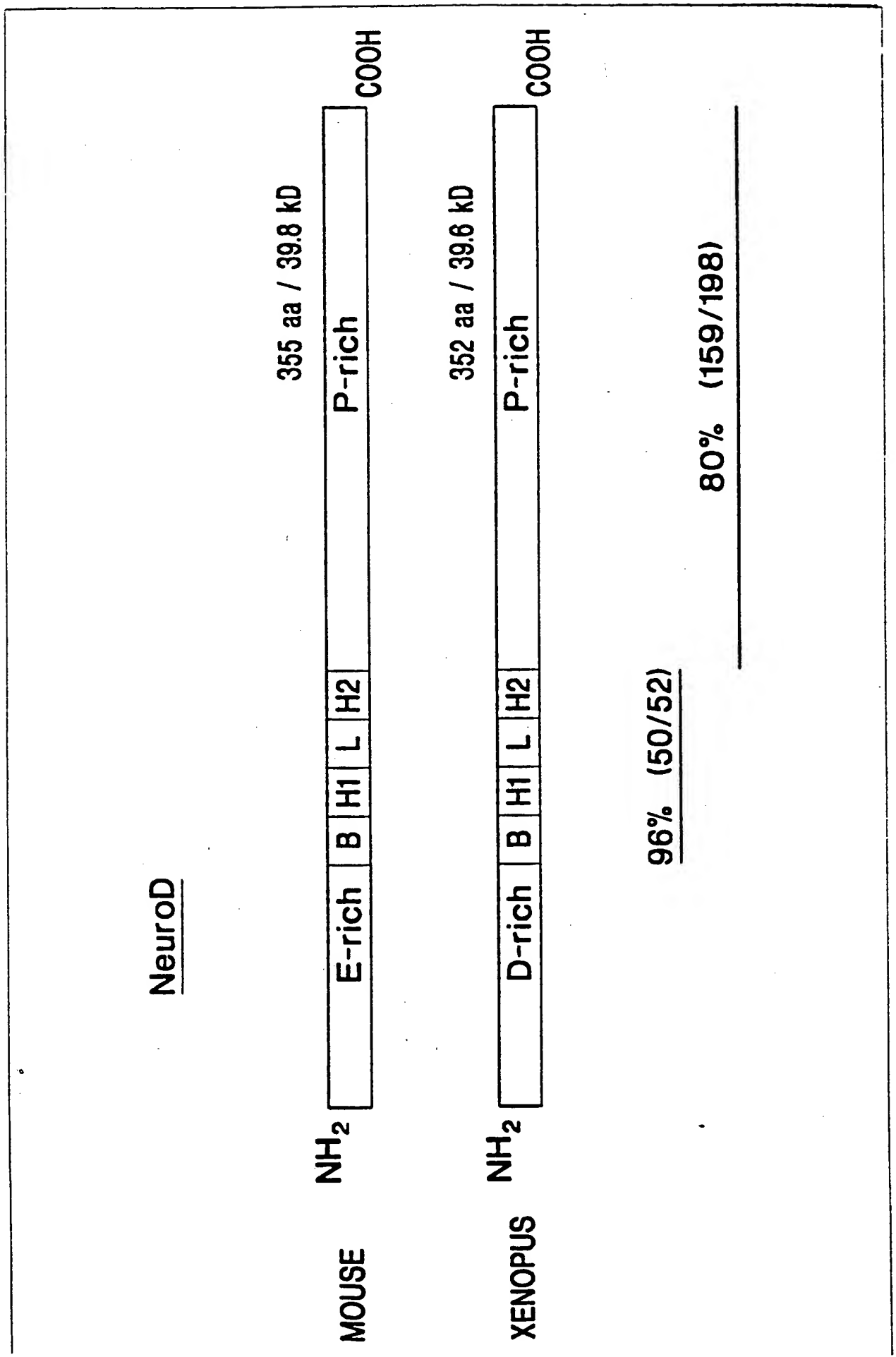
7. An antibody or antigen-binding fragment thereof that binds to the recombinant peptide of claim 5.

8. An antibody or antigen-binding fragment thereof that binds to the recombinant peptide of claim 6.

9. An antibody or antigen-binding fragment thereof that binds to a peptide selected from among SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:9, and SEQ ID NO:11.

10. An antibody or antigen-binding fragment thereof that binds to a peptide selected from among amino acid residues 117-156 of SEQ ID NO:2, amino acid residues 118-157 of SEQ ID NO:4, amino acid residues 117-156 of SEQ ID NO:9, and amino acid residues of 91-130 of SEQ ID NO:11.

11. A method for inducing differentiation of a non-neuronal cell into a neuron, comprising introducing a nucleic acid molecule of claim 1 into the non-neuronal cell.


NeuroD
MOUSE
NH2
E-rich
B
H1
L
H2
P-rich
COOH
355 aa / 39.8 kD
XENOPUS
NH2
D-rich
B
H1
L
H2
P-rich
COOH
352 aa / 39.6 kD
96% (50/52)
80% (159/198)

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.

US CL :Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : Please See Extra Sheet.

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, DIALOG: MEDLINE, BIOSIS, Derwent Biotechnology Abstracts, CAS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A, P	US, A, 5,322,801 (KINGSTON ET AL.) 21 JUNE 1994, see entire document.	1-11

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

20 JUNE 1995

Date of mailing of the international search report

05 JUL 1995

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

John L. LeGUYADER

Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

C07K 14/435, 14/47, 16/00; C12N 5/16, 15/12, 15/85

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

435/69.1, 172.3, 240.2, 320.1; 536/23.4, 23.5, 24.33; 530/350, 387.1

B. FIELDS SEARCHED
Minimum documentation searched
Classification System: U.S.

435/69.1, 172.3, 240.2, 320.1; 536/23.4, 23.5, 24.33; 530/350, 387.1